Oral human papillomavirus infection incidence and clearance: a systematic review of the literature

Zoe C. Wood, Christopher J. Bain, David D. Smith, David C. Whiteman and Annika Antonsson*

Abstract
Subclinical oral human papillomavirus (HPV) infection that persists for decades is likely to precede an HPV-driven squamous cell carcinoma of the head and neck, but little is known about the natural history of oral HPV. We systematically reviewed and abstracted data from nine manuscripts that examined human immunodeficiency virus-negative and cancer-free subjects for oral HPV DNA to determine the pooled baseline prevalence and incidence of newly acquired oral HPV infections, and specifically for HPV-16. We also documented the clearance rate and the median time to clearance, where data existed. Of 3762 individuals, 7.5% had an oral infection with any HPV type (1.6% for HPV-16). Meta-regression analysis estimated the 12-month cumulative incidence to be 4.8% (95% confidence interval 3.2–7.3%). The overall oral HPV clearance was reported to be 0–80% between studies, and the median time to clearance from 6.5 to 18 months. Oral HPV-16 clearance was 43–83%, and median time to clearance for HPV-16 was 7–22 months. Oral HPV prevalence, incidence and clearance vary considerably between published studies from different geographical regions. Further research is required to identify predictors of persistent oral HPV infection. Measurable baseline prevalence was observed in all studies, as well as non-trivial incidence of newly acquired oral HPV infections and incomplete clearance.

INTRODUCTION
Mucosal high-risk human papillomaviruses (HPVs) have been identified as a necessary cause of cervical cancers [1] and have been detected in 50–70% of genital and anal cancers [2]. There is increasing evidence linking HPV infections of the oral mucosa to mucosal squamous cell carcinomas (SCCs) of the head and neck (hereafter HNSCCs) [3, 4]. HNSCC is the sixth most common type of cancer worldwide with an estimated 575,000 new cases diagnosed annually. There has been an increase in incidence recently in younger patients with minimal levels of exposure to the classical risk factors of smoking and alcohol [3, 5], which have been shown to be related to HPV infection, especially for cancer of the oropharynx, presumably acquired via increased oral exposure to infected anogenital sites with changing sexual behaviours [6–9]. Relatively little is known about the natural history of oral HPV infection, but it is likely that a subclinical oral HPV infection that persists for decades precedes the development of HPV-driven HNSCC (as has been established for cervical cancer [1]). To further understand the dynamics of this infection we review here the existing literature to estimate the incidence and clearance of HPV DNA detected in oral specimens collected from cancer-free, human immunodeficiency virus (HIV)-negative individuals.

RESULTS
Nine publications comprising 3762 cancer-free participants were identified and are presented in Table 1 [10–18]. Sample sizes ranged from 131 [11] to 1626 [12] participants. The studies were performed in the USA, Brazil, Mexico and Finland. The HPV Infection in Men (HIM) Study investigated men only from three countries (USA, Brazil and Mexico) [12, 14]. The Finnish Family HPV Study analysed oral specimens from pregnant women, fathers-to-be and mothers [11, 13, 15, 16]. Beachler et al. included a US high-risk population (defined as ‘HIV-uninfected people at risk of HIV infection’) in their recent study [10]. Two studies [18, 19] analysed oral HPV prevalence and incidence in US university students.

Oral specimen collection and DNA extraction methods varied between studies, but most studies used MY09/11 PCR (alone or in combination with GP5+/6+) for HPV detection (Table 2). Time between follow-up varied between studies,
with 4-, 6- or 12-monthly intervals between samples being used (Table 2).

**Oral HPV prevalence**

Overall, the prevalence estimates were heterogeneous across studies, ranging from a low of 2.5 % in a study of US university students to as high as 20 % in a US high-risk population [12, 18]; the average baseline oral HPV prevalence across studies was 7.5 % (284/3762; Table 3). Most of the studies presented HPV-16 type-specific prevalence, which ranged from 0.2 % [18] to 17 % [13]. The average HPV-16 prevalence across studies was 1.6 % (52/3162; Table 3). Oral HPV prevalence for each study is reported in Table 3.

**Oral HPV incidence**

The incidence of new oral HPV infections (all types, and HPV-16 specifically) was even more heterogeneous between studies than reported for oral HPV prevalence (see Table 3). The lowest overall oral HPV incidence (2 %) was reported in US university students [18] and the highest (69 %) in fathers-to-be in the Finnish study [11]. The type-specific HPV-16 incidence ranged from 0.6 % in the HIM Study [12] to 65 % in the Finnish Family HPV Study [13].

Using meta-regression analysis under a fixed effects assumption, we estimated the 12-month cumulative incidence of oral HPV infection to be 4.8 % [95 % confidence interval (CI) 3.2 – 7.3 %] across the six studies. When we assumed random effects rather than fixed effects, we detected significant heterogeneity (Q test \( P < 0.001 \)); however, the 12-month incidence estimate was similar at 5.0 % (95 % CI 3.0 – 8.3 %; Fig. 1).

**Oral HPV clearance**

Clearance for any oral HPV infection spanned from 0 % among pregnant women in the Finnish study [16] to 80 % in the US high-risk population [10], and time to clearance of any HPV type ranged from 6 to 18 months [11, 15]. HPV-16 type-specific clearance was less heterogeneous than overall HPV clearance and was lowest (43 %) in Finnish mothers [13] and highest (83 %) in two separate studies from the USA (the HIM Study and the high-risk population) [10, 12]. Time to clearance of HPV-16 took 7 months in the HIM Study [12] and around 20 months for participants in three publications from the Finnish Family HPV Study [11, 13, 15]. Oral HPV clearance (as percentage cleared) was reported in three of the studies and time to clearance in two (Table 3).

**DISCUSSION**

HPV-driven HNSCCs have been increasing in recent decades [8, 20], so it is surprising that little is known about the fundamental dynamics of oral HPV infection, including basic epidemiological constructs such as the prevalence of infection, the incidence of new infections and the rate of clearance of existing infections. Such knowledge about the natural history of oral HPV infection would help to better understand the likely trajectory of HPV-driven HNSCCs in
<table>
<thead>
<tr>
<th>Study</th>
<th>Reference</th>
<th>Oral specimen collection method</th>
<th>Specimen processing/DNA extraction</th>
<th>HPV detection method (prevalence/incidence)</th>
<th>Number of HPV genotypes targeted (LR, low risk; HR, high risk)</th>
<th>Follow-up visits (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The HPV Infection in Men (HIM) Study</td>
<td>[12]</td>
<td>Rinse-and-gargle</td>
<td>Robotic MDx Media Kit</td>
<td>PGMY09/11 Li consensus primer system + Linear Array (Roche)</td>
<td>37 (LR: 6, 11, 26, 40, 42, 53–55, 61, 62, 64, 66, 67, 69, 70–73, 81–84, CP6109, IS39; HR: 61, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 68)</td>
<td>6, 12, 18, 24, 30, 36, 42, 48</td>
</tr>
<tr>
<td></td>
<td>[14]</td>
<td>Rinse-and-gargle</td>
<td>QIAamp Media MDx Kit</td>
<td>PGMY09/11 PCR; Linear Array (Roche) and INNO-LiPA HPV Genotyping Extra</td>
<td>1 (HR: 16)</td>
<td>6, 12, 18, 24, 30, 36, 42, 48</td>
</tr>
<tr>
<td></td>
<td>[13]</td>
<td>Scrapings</td>
<td>High salt method</td>
<td>MY09/11 and GP5+/6+</td>
<td>As above</td>
<td>2, 6, 12, 24, 36, 72</td>
</tr>
<tr>
<td></td>
<td>[15]</td>
<td>Scrapings</td>
<td>High salt method</td>
<td>MY09/11 and GP5+/6+</td>
<td>As above</td>
<td>2, 6, 12, 24, 36, 72</td>
</tr>
<tr>
<td></td>
<td>[16]</td>
<td>Brush of oral mucosa</td>
<td>Protease K</td>
<td>MY09/11 and GP5+/6+</td>
<td>12 (HR only: 16, 18, 31, 33, 35, 39, 45, 51, 52, 54, 56, 58)</td>
<td>2, 6, 12, 24</td>
</tr>
<tr>
<td>The Persistent Oral Human Papillomavirus Study (POPS)</td>
<td>[10]</td>
<td>Rinse-and-gargle</td>
<td>QIAsymphony SP</td>
<td>PGMY09/11 PCR and Linear Array HPV Genotyping Test (Roche)</td>
<td>37 (LR: 6, 11, 26, 40, 42, 53–55, 61, 62, 64, 66, 67, 69–72, 81–84, 89 (CP6108), IS39; HR: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73)</td>
<td>6, 12, 18, 24, 30, 36</td>
</tr>
</tbody>
</table>
the future. With this systematic review we wanted to estimate, in particular, oral HPV incidence and clearance, and also oral HPV prevalence in published studies.

The overall oral HPV prevalence ranged from 2.4 to 20 %, with a pooled prevalence of 7.5 % across the different studies. HPV-16 was the most common HPV type identified in the studies with an oral prevalence ranging from 1 to 17 %. The oral HPV prevalence was consistently high in the four reports from the Finnish Family HPV Study [11, 13, 15, 16], ranging from 17 to 18 %. However, the highest oral HPV prevalence, 20 %, was found in high-risk (at-risk, but HIV-uninfected) persons. This oral HPV prevalence is higher than what was found in a recent systematic review of men who have sex with men (9 % HIV– and 16 % HIV+ individuals) who are considered high-risk [21].

Previous reports of oral HPV prevalence vary greatly between different geographical areas. The majority of other publications on the prevalence of oral HPV infection (without data on incidence) have been performed in the USA [17, 22–24], along with several from other geographical regions [15, 25, 26]. The largest (n=5579, a cross-sectional study in the USA conducted as part of the National Health and Nutrition Examination Survey) reported an overall oral HPV prevalence of 6.9 % [22], with a significant difference in oral HPV prevalence between men (10.1 %) and women (3.6 %; P<0.001).

We found that incidence and clearance varied considerably between published studies of different sizes and from various geographical regions. The estimated 12-month incidence from all studies was 4.8 % (95 % CI 3.2–7.3 %). Oral incidence varied widely between different populations, and was lowest in the HIM Study [12] and was high in both US high-risk individuals [10] as well as fathers-to-be in the Finnish Family HPV Study [11]. Oral HPV clearance data were limited and discrepant across studies (0–80 %), with the lowest clearance reported in the Finnish study [16], and the highest rate of clearance in a US high-risk population [10].

There are difficulties in comparing various published data on the natural history of HPV as there are wide differences in findings and study methods. There are several methodological reasons that might explain the differences in the reported oral HPV prevalence, incidence and clearance, which include different sampling methods used (mouthwash, saline and cytobrush), method of DNA extraction and HPV detection methods used [different PCR primer pairs and methods for HPV typing (sequencing versus arrays)]. All US studies (n=4) included here collected their samples with a rinse and gargle method, while the Finnish study used scrapings or brushing of oral mucosa. In a previous study [27], we collected both saline mouthwash samples and swabs from the tonsillar fossa in the same HNSCC patients, and found that the mouthwash samples had significantly higher HPV prevalence than the swab samples (25 versus 9 %).

D’Souza et al. [28] compared five different methods of DNA extraction of 20 HIV-positive participants using Scope mouthwash samples and MY PCR by reverse dot blot hybridization, and found large differences between the methods. The method of DNA extraction impacted not only DNA quality (measured with β-globin PCR, range 50–100 % positive), but also the ability to determine HPV status (range 11–65 %) and number of HPV types (2–13 HPV types) [28].

A recent publication [29] compared the performance of the commonly used HPV detection and typing methods GP5+/6+ PCR followed by reverse dot blot hybridization (GP5+/6+ RLB) and multiplex type-specific E7-based bead-based multiple HPV genotyping (MPG) assay (E7-MPG) on almost 5000 cervical cell and urine samples. Both methods detected 22 % HPV-positive samples, and 14 % were positive with the E7-MPG method only and 0.5 % positive with the GP5+/6+ RLB method only. Nested and GP5+/6+ PCR was used by all studies included in our analysis, but this method of HPV detection was not evaluated [29]. The nested MY09/11 and GP5+/6+ PCR has been found to be more type-sensitive, detecting a wider range of HPV types, low-copy-number HPV infections and more infections with multiple HPV types than MY09/11 PCR or GP5+/6+ PCR only [30]. To our knowledge, there are no studies that have compared E7-MPG with nested MY09/11 and GP5+/6+ PCR. Thus, there is merit to standardizing methods of sample collection, DNA extraction and HPV detection to ensure comparability between studies in the future.

Our systematic review of the existing literature faced a number of challenges. Firstly, there were very few studies that have reported on oral HPV incidence and clearance. Secondly, those few studies have used different methods, both in regards to collecting samples and to analysing them for HPV. Lastly, the heterogeneity between the available studies, including geographical area, sample size, gender restrictions and different durations of follow-up, severely limited our ability to generalize across studies.

The natural history of oral HPV infection is not known, nor whether an oral HPV infection can progress to HNSCC. It is clear, however, that HPV-positive oropharyngeal SCCs, in particular tonsillar SCCs, which have the highest HPV prevalence, are increasing. Based on what is known about HPV-driven cervical cancer, one might predict that HPV-driven HNSCCs will also be preceded by subclinical oral HPV infections that persist for decades. However, due to currently inadequate sampling methods, we do not know whether HPV-positive precursor lesions exist, as HPV-positive precursor lesions have not been identified in the oral cavity or oropharynx. As cervical cytology screening (Pap-test) has been very successful in preventing cervical cancers, a similar cytology test might be useful for early detection of oropharyngeal SCCs. To our knowledge, only one study has investigated this [31] and found that an oropharyngeal Pap-test probably will not pick up pre-lesions. This finding could
### Table 3. Prevalence, incidence and clearance of studies that investigated oral HPV infections in cancer-free, HIV-negative populations; n total=3762

<table>
<thead>
<tr>
<th>Study Reference</th>
<th>Study population</th>
<th>Any HPV baseline prevalence [n (%)]</th>
<th>HPV-16 baseline prevalence [n (%)]</th>
<th>Time period for incidence (months)</th>
<th>Incidence for any HPV [n (%)]</th>
<th>Incidence for HPV-16 [n (%)]</th>
<th>Any HPV cleared infections †</th>
<th>Any HPV median time to clearance (months)</th>
<th>HPV-16 infections cleared [n (%)] ‡</th>
<th>HPV-16 median time to clearance (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The HPV Infection in Men (HIM) Study [12]</td>
<td>Men</td>
<td>71 (4.0)</td>
<td>10 (1.0)</td>
<td>12</td>
<td>12</td>
<td>115 (4.4)</td>
<td>18 (0.6)</td>
<td>45 (55 %)</td>
<td>6.9</td>
<td>5 (83 %)</td>
</tr>
<tr>
<td>The Finnish Family HPV Study [11]</td>
<td>Fathers-to-be</td>
<td>24 (18.3)</td>
<td>36</td>
<td>90 (60.2)</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
<td>30 (23 %)</td>
<td>90 (62 %)</td>
</tr>
<tr>
<td></td>
<td>Mothers</td>
<td>56 (100)</td>
<td>55 (96.4 %)</td>
<td>6.5</td>
<td>18 (58.3 %)</td>
<td>21.7 (mean)</td>
<td>471 (63 %)</td>
<td>207 (24 %)</td>
<td>95 % CI 23-81.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Women and men</td>
<td>56 (100)</td>
<td>34 (60.7 %)</td>
<td>6.5</td>
<td>18 (58.3 %)</td>
<td>21.7 (mean)</td>
<td>471 (63 %)</td>
<td>207 (24 %)</td>
<td>95 % CI 23-81.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not specified</td>
<td>18</td>
<td>18.6 (mean)</td>
<td>0 (0 %)</td>
<td>0 (0)</td>
<td>18</td>
<td>20.7 (mean)</td>
<td>32</td>
<td>4.8 % (95 % CI 3.2–7.3 %)</td>
<td></td>
</tr>
<tr>
<td>The Persistent Oral Human Papillomavirus Study (POPS) [10]</td>
<td>At-risk HIV-uninfected persons</td>
<td>34 (2.4)</td>
<td>2 (0.6)</td>
<td>3</td>
<td>3</td>
<td>523 (62 %)</td>
<td>20 (2.0)</td>
<td>0 (0)</td>
<td>3.2–7.3 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>University students</td>
<td>284 (76 %)</td>
<td>523 (62 %)</td>
<td>3.2–7.3 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Time period for observation varies across studies (presented in column 6). †Excluding prevalent infections. ‡Reported as 12 months for all studies except the study by Pierce Campbell who reported clearance over 44 months. §Pooled oral HPV prevalence from the studies of Kreimer, Kero, Rautava, Beachler, Edelstein and Pickard. From the Finnish Family HPV Study, only data from Kero and Rautava were included to avoid overlap. ¶Twelve-month incidence estimate from all studies reporting incidence using a meta-regression model.
reflect the difficulties in sampling the tonsillar crypts. Given that oral HPV prevalence is low and that HPV-driven HNSCC is a rare type of disease, very large studies are needed. Because oral HPV infection is relatively uncommon, larger studies are needed to ensure sufficient statistical power, in particular to identify factors associated with persistence and clearance.

In conclusion, we observed a measurable baseline prevalence in all studies, as well as non-trivial incidence of newly acquired oral HPV infections and incomplete clearance, despite differences in methodologies and approaches. The non-trivial incidence and incomplete clearance of oral HPV infections are potentially worrisome findings, as these data would suggest that the pool of people carrying oral HPV infections may be increasing. Given the lack of knowledge of the natural history of oral HPV infection, new prospective studies that focus on oral HPV incidence and clearance are needed. Because new episodes of oral HPV infection are relatively uncommon in any given population, studies with large sample sizes are needed to ensure sufficient statistical power to capture these ‘rare events’, all the more so when attempting to identify factors associated with persistence and clearance among the subset of participants developing new infections. Such knowledge is fundamental to understanding the likely trajectory of oropharyngeal cancer in the future and to predict the impact that HPV vaccination will have on oral HPV infection and HPV-driven oropharyngeal cancers in the future.

METHODS

Study selection

We searched the National Institutes of Health 'PubMed' search engine for citations published within a 10-year period (2006–2016) using the medical subject heading terms 'Papillomavirus' and 'Oral' as well as the terms 'Incidence' or 'Clearance' in all fields. The search was limited to English publications in humans. By using these parameters, we identified 84 potentially eligible manuscripts. After reviewing their titles and abstracts for relevance, we identified 28 studies that appeared to evaluate the incidence and/or clearance of HPV DNA in oral specimens collected from healthy individuals (i.e. cancer-free and HIV-negative). From this selection, we then excluded studies that focused on individuals with precancerous lesions/papillomatosis (n=6) and studies of immunosuppressed populations (HIV-positive) (n=3). However, if a study included HIV-negative and HIV-positive individuals and HPV data were provided stratified by HIV status, we utilized the available data. We excluded commentaries and other systematic reviews (n=3) and studies that focused exclusively on newborns and children (n=8). Using these criteria, eight eligible manuscripts remained for analysis. After carefully examining related
citations and references of these papers, another publication was included; all nine used DNA-based testing methods on oral specimens and had >50 participants. Some of the papers belonged to overarching studies. Specifically, two manuscripts reported the ‘HPV Infection in Men (HIM) Study’ and four manuscripts came from the ‘Finnish Family HPV Study’ (FFHPVS). Thus, the nine publications were derived from a total of five studies. All studies obtained informed consent from their respective participants.

Data abstraction

Data were abstracted (by Z. C. W.) and checked (by A. A.) on the following variables: first author, year of data collection, country from which the sample was obtained, study population (i.e. university students, pregnant women), number of samples with good quality DNA (β-globin positive by PCR), age (mean or median and range), sex distribution, oral specimen collection method, specimen processing method/DNA extraction, HPV detection method, HPV genotypes targeted, overall prevalence of HPV at baseline, type-specific prevalence of HPV-16 at baseline, HPV incidence, type-specific incidence for HPV-16, time period for overall oral HPV incidence and for HPV-16, clearance of at least one prevalent HPV type and median time to clearance.

Measures

Oral HPV prevalence was reported as the proportion of all samples that tested positive for any type of HPV; we also recorded HPV-16-specific prevalence where available. Because of within-study overlap of samples (i.e. participant samples reported in more than one publication) we used data from Kreimer et al. [12] for the HIM Study and data from Kero et al. [11] (fathers-to-be; n=131) and Rautava et al. [15] (mothers; n=324) for the Finnish Family HPV Study, to calculate the oral HPV prevalence (overall and for HPV-16 separately).

All studies that measured incidence used the same definition for oral HPV incidence – that is, participants negative for oral HPV infection at baseline were included, and the first HPV-positive sample was defined as an ‘incident’ or ‘newly acquired’ infection. Person-time for a newly acquired HPV infection was calculated from the baseline sample to the first HPV-positive sample, with the assumption that a new infection arose on the date the HPV-positive sample was collected. Oral HPV incidence (overall and for HPV-16) was measured differently between studies: one study reported 3-month incidence [18], two reported 12-month incidence [12, 19], two reported 24-month incidence [10, 16] and one reported 36-month incidence [11]. To estimate the 12-month incidence for all studies, we fitted two weighted meta-regressions, one with fixed effects and one with random effects, with weights inversely proportional to the standard errors of the reported incidence rates. After taking a log-transformation of the incidence rates, we fitted time as a linear independent predictor for log (incidence). The predicted incidence rates and standard errors at each time point were transformed back to the linear scale and these served as our incidence estimates for a given time point.

Clearance was only measured by three studies, the HIM Study, the Finnish Family Study and the US study of a high-risk population. In the HIM Study, HPV type-specific clearance was defined as an HPV-positive participant testing HPV-negative after having tested positive previously and excluding prevalent infections and those who tested positive for the first time in their last sample [12]. Time to clearance was estimated by the Kaplan–Meier method. For women in the Finnish Family HPV Study, clearance was defined as an event (at any follow-up visit) when a previously HPV-positive test turned out to be negative and remained HPV-negative until the end of the follow-up [13]. For men in the Finnish Family HPV Study, baseline HPV-positive cases that cleared the infection during the follow-up was defined as events of clearance [11]. In the high-risk population study [10], clearance was defined as testing negative once after a positive test.

Funding information

D.C.W (APP1058522) and A.A (APP1065293) were funded by Research Fellowships from the NHMRC. The funding bodies played no role in the design or conduct of the study. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the funding bodies.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Ethical statement

Ethics approval not obtained as this is a systematic review of the literature.

References


---

**Five reasons to publish your next article with a Microbiology Society journal**

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.