



HAL
open science

Incidence and duration of human papillomavirus infections in young women: insights from a bimonthly follow-up cohort

Thomas Bénéteau, Soraya Groc, Carmen Lía Murall, Vanina Boué, Baptiste Elie, Nicolas Tessandier, Claire Bernat, Marine Bonneau, Vincent Foulongne, Christelle Graf, et al.

► To cite this version:

Thomas Bénéteau, Soraya Groc, Carmen Lía Murall, Vanina Boué, Baptiste Elie, et al.. Incidence and duration of human papillomavirus infections in young women: insights from a bimonthly follow-up cohort. *Infectious diseases (London, England)*, 2024, pp.1-10. 10.1080/23744235.2024.2427223 . hal-04822236

HAL Id: hal-04822236

<https://hal.science/hal-04822236v1>

Submitted on 6 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Incidence and duration of human papillomavirus infections in young women: insights from a bimonthly follow-up cohort

Thomas Bénéteau^{1,*}, Soraya Groc², Carmen Lía Murall³, Vanina Boué¹, Baptiste Elie¹, Nicolas Tessandier⁴, Claire Bernat^{1,5}, Marine Bonneau⁶, Vincent Foulongne², Christelle Graf⁶, Sophie Grasset¹, Massilva Rahmoun¹, Michel Segondy², Vincent Tribout⁷, Jacques Reynes⁸, Christian Selinger^{1,9}, Nathalie Boulle⁶, Ignacio G. Bravo¹, Mircea T. Sofonea^{2,10}, Samuel Alizon^{1,4,*}

¹ MIVEGEC, Univ Montpellier, CNRS, IRD, France

² PCCEI, Univ Montpellier, INSERM, EFS, Univ Antilles, Montpellier, France

³ Department of Biological Sciences, Université de Montréal, Montréal, Canada

⁴ Center for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, Université PSL, Paris, France

⁵ Institut de Génomique Fonctionnelle, Université de Montpellier, CNRS, INSERM, Montpellier, France

⁶ Department of Obstetrics and Gynaecology, Centre Hospitalier Universitaire de Montpellier, Montpellier, France

⁷ Center for Free Information, Screening and Diagnosis (CeGIDD), Centre Hospitalier Universitaire de Montpellier, Montpellier, France

⁸ Department of Infectious and Tropical Diseases, Centre Hospitalier Universitaire de Montpellier, Montpellier, France

⁹ Swiss Tropical and Public Health Institute, Basel, Switzerland

¹⁰ Department of Anesthesiology, Critical Care, Intensive Care, Pain and Emergency Medicine, CHU Nîmes, Nîmes, France

* Corresponding authors: thomas.beneteau@ird.fr and samuel.alizon@cnrs.fr

Abstract

Background. We studied the duration of HPV detection and risk of (re-) detection for 25 HPV genotypes in a cohort of 132 women followed every eight weeks for up to two years between 2016 and 2020. Participants were between 18 and 25 years old at inclusion and half of them were vaccinated against HPV. They were recruited near the University and the STI detection centre in Montpellier, France.

Methods. We used genotype-specific longitudinal data to characterise the dynamics of HPV-detected episodes. We investigated the contribution of viral and host factors to the variations in the duration of HPV detection, and the time before (re-)detection of the same genotype using multivariate Cox regression models with frailty at the patient level.

Findings. We detected at least one HPV episode in 74% of the participants and re-detected the same genotype in 47% of them. Covariates related to socio-economic difficulties were associated with a lower risk of detectability loss (hazard ratio 0.45 with a 95% confidence interval, CI, from 0.21 to 0.97). The number of lifetime sexual partners was strongly associated with an increased risk of new positive detection (hazard ratio 2.40 with a 95%CI from 1.07 to 5.39). In contrast, vaccination was associated with a lower risk of displaying incident infections (hazard ratio of 0.64 with a 95%CI from 0.43 to 0.96).

Conclusion. In the short term, vaccination shows clear signs of protection against new HPV detections, including for some genotypes not targeted by the vaccine, such as HPV31 and HPV51.

Keywords

Papillomaviruses ; Vaccination ; Persistence ; Public Health ; Clearance

Introduction

Among the hundreds of Human Papillomavirus (HPV) genotypes identified [1], a handful is responsible for the development of multiple cancers accounting for more than 342,000 deaths worldwide each year [2]. They are also one of the most common sexually transmitted infections (STIs), with models suggesting that, by the age of 45, more than 80% of the population has already been infected by HPV [3]. Initial infection is thought to occur during the first sexual intercourse, with the prevalence peaking after sexual debut [4], and the risk of contracting a new HPV infection increases with the number of sexual partners [5]. Some occurrences of vertical HPV transmission have also been reported, both *in utero* or through post-natal acquisitions [6], but there is little evidence these infections could persist and still be detectable dozens of years later. Most HPV infections are asymptomatic and usually not detectable after three years of follow-up [7]–[9], which is largely attributed to infection clearance [10]. However, the interpretation of re-detection events is delicate as they might be caused by at least three distinct mechanisms such as true new infection, transient deposition from a sexual partner, or reactivation of a subclinical infection [9], [11]. HPV detection is likely a combination of these scenarios and being able to distinguish between them remains a major challenge to optimise public health policies, evaluate the effectiveness of catch-up vaccination, or organise the age stratification of vaccination policies.

Most of our knowledge on the natural history of HPV infections originates from longitudinal clinical studies, especially the control arm of vaccine clinical trials [7], [8] but also additional cohorts [4]. In this work, we analysed data from the PAPCLEAR cohort [12], which stands out in terms of the frequency of the sampling. Contrary to vaccine trials in which on-site visits typically occurred every 6 months or more, here, infected participants visited the clinic every 8 weeks and were screened for 25 HPV genotypes.

We analysed the PAPCLEAR follow-up data using survival models with frailty effects at the patient level and accounting for the censoring of the data. The main goal of this work was to better characterise the dynamics of cervical HPV infections in healthy young women, both in terms of the duration of positive genotype-specific HPV episodes and of the duration between consecutive positive HPV detection for the

same genotype. In particular, we evaluated the contributions of viral and host factors on these quantities using Cox regression models.

Materials and methods

Study design and participants

The PAPCLEAR cohort is a monocentric prospective longitudinal study following 149 women (i.e. assigned female at birth). Figure 1 summarises the PAPCLEAR protocol and further details can be found in [12]. In short, the study was run between 2016 and 2020 at the STI detection centre (the CeGIDD) of the University Hospital of Montpellier, France. It was advertised through posters and leaflets handed out in the University area. The inclusion criteria were that participants had to be aged between 18 and 25 years old at inclusion, live in the area of Montpellier (France), and report having at least one new sexual partner over the last 12 months. Women with a history of HPV-associated pathology and women planning a pregnancy within the first year of inclusion, or already pregnant, were excluded from the study. These criteria were set to maximise the number of incident HPV infections. Eligible participants were invited for a baseline visit during which, after official enrolment, biological samples were collected and they filled in an extensive questionnaire related to demographic, socioeconomic, and behavioural risk factors. Follow-up on-site visits were scheduled 4 weeks later for the results visit and then 8 weeks later for infected participants (HPV⁺ arm) and 16 weeks later for uninfected ones (HPV⁻ arm). During each on-site visit, a gynaecologist or a midwife performed a general exam during which several samples were collected in the genital tract only, including cervical smears that were used to screen for HPVs. Participants also filled in follow-up questionnaires at each of these visits to notify changes in their habits. All infected participants were followed for up to 2 years and participants who stayed in the HPV⁻ arm were followed for at most 8 months. All participants provided written informed consent.

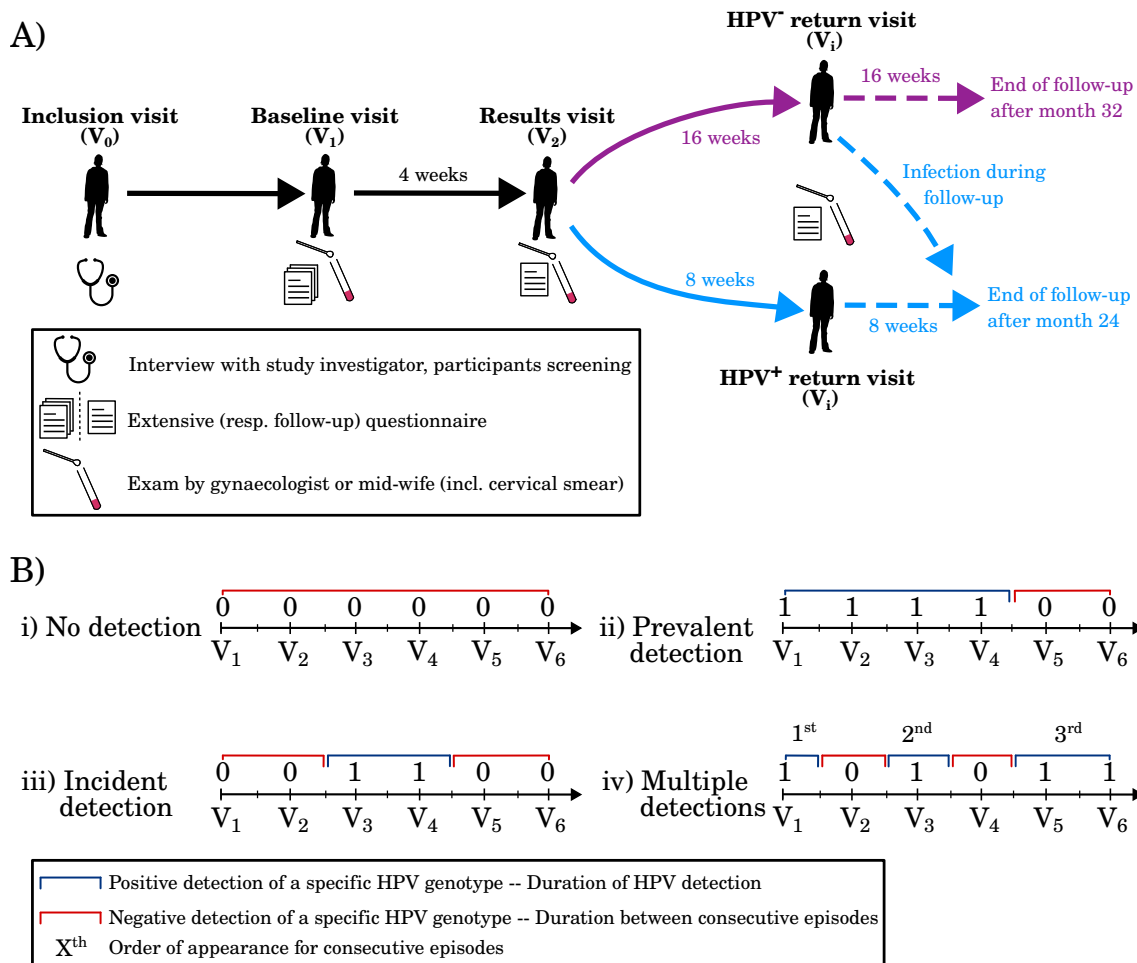


Figure 1: **General structure of the PAPCLEAR study and main HPV detection patterns.** A) Overview of the general structure of the PAPCLEAR study. During each on-site visit (excluding V_0), a gynaecologist or midwife performed a general exam during which several samples were collected, including a cervical smear, and participants filled in a questionnaire related to demographic, socioeconomic, and behavioural risk factors. The purple arrows refer to the HPV⁻ arm and the blue arrows to the HPV⁺ arm. B) Diagrams recapitulating the various scenarios encountered in the analysis of cervical samples. For each visit, a ‘1’ indicates a genotype-specific positive HPV detection and a ‘0’ genotype-specific negative HPV detection. The blue intervals were used in the analysis of the duration of HPV detection and the red ones were used in the analysis of the time between consecutive episodes.

Genotyping

We first tested for the presence of alpha-papillomaviruses in the cervical smears using the DEIA assay [13]. DEIA-positive samples were genotyped using the LiPA25 assay, which was chosen for its sensitivity and can detect up to 25 different HPV genotypes [14]. Among these, we refer to high-risk (HR) genotypes for HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68 [15], and to low-risk (LR) for the remaining 12 (HPV6, 11, 34, 40, 42, 43, 44, 53, 54, 70, and 74). For samples that were LiPA25-negative

and DEIA-positive, the genotype was determined using the PGMY PCR amplification [16] and Sanger sequencing of the PCR product. In some cases, the genotype identification was impossible, most likely because of a co-infection by multiple HPV genotypes amplified by the PGMY PCR, in which case the sample was labeled as ‘non-typable’. DEIA-negative samples were not analysed further.

Analysis inclusion criteria

We excluded the data from all the women with less than three visits, which amounted to 17 participants (4 were seen once, and 13 were seen only twice). The unit of observation was the HPV genotype, meaning that each participant could contribute to multiple observations. Following earlier studies, we assumed the dynamics of each genotype to be independent at the participant level and between participants [17]. Unless specified otherwise, we did not assume any genotype-specific effects.

Events definition

In subsequent analyses, HPV-genotype-specific detection at the patient level serves as the primary unit of observation. Following previous studies [18], an HPV episode was defined as ‘prevalent’ if detected at inclusion and as ‘incident’ if the genotype was detected at posterior visits but not at inclusion. During a follow-up, several distinct episodes were sometimes recorded for a given HPV genotype. These episodes were distinguished according to their order of appearance (first, second, third) in the following analyses (Figure 1B). Patterns of positive detection separated by only one negative visit, sometimes referred to as ‘intermittent detection’ [18], were handled differently between studies given the absence of consensus in the literature regarding the way to deal with such data. In our analysis, we assumed that two such episodes were separate events.

For each genotype and each participant, we analysed the duration of HPV detection and the duration between HPV-positive episodes. Survival functions for these quantities were computed using the non-parametric Nelson-Aalen estimator [19]–[21].

For a given episode, we defined the duration of HPV detection as the duration between the midpoint

at the start of the episode and the midpoint at the end of the episode. We included all incident episodes, even the shortest episodes that were only detected during one visit. Similarly, the time between two consecutive positive episodes was computed as the duration from the midpoint at the end of the expired episode (or from inclusion for the first incident episodes) to the midpoint at the start of the new episode. Extensive information can be found in Supplementary Methods A2 and in Figure 1B.

Statistical analyses

We investigated differences in patterns of detection between HR-genotypes and LR-genotypes using log-rank tests [22], [23]. To test for differences between two populations, we used Fisher’s exact test for qualitative variables and Wilcoxon-Mann-Whitney’s test for quantitative variables. We display the raw p -values in the results and use a 5% threshold for statistical significance.

We used Cox proportional hazards models [24] with frailty at the participant level [25] to infer the contribution of behavioural and socio-economic factors on both times to incident HPV episodes and duration of HPV detection. We stratified the Cox regression with different baseline hazard functions according to the order of occurrence of the episodes (first, second, or third detection). We assumed no interaction between the strata and conducted multivariate analyses using multi-model inference [26], [27]. Model selection was performed using the corrected Akaike Information Criterion (AICc) as a metric for the penalised goodness of fit [28]. We evaluated the goodness of fit for all sub-models of the maximum model, i.e. the one with all the covariates, and estimated the hazard ratios of the Cox regression using a full averaging procedure on the best models (i.e. those with an AICc no larger than the minimum AICc+2). More details regarding our statistical methodology can be found in the Supplementary Materials A3.

We included the following covariates in the analysis: the number of lifetime sexual partners (LTSP), the body mass index (BMI) at baseline, the self-declared ethnic origin (Caucasian vs. non-Caucasian or mixed-origin), the HPV vaccination status, the sexual preferences, the use of condoms or oral hormonal contraception, an indicator of financial difficulties (participants who declined medical care because of

financial reasons in the 12 months prior to inclusion), the number of years between inclusion and first menstruation, the number of years between inclusion and first sexual intercourse and the smoking status (past, current or never). Additional details about the covariates' values can be found in Table S1 in the Supplementary Materials.

All statistical analyses were conducted using R 4.2.2 with additional packages listed in Supplementary Materials A7.

Results

Overview of the study population

For the 132 women included in the analysis, the follow-up duration was on average 309.5 days, median with an IQR of 203 to 467.5. There was no significant difference in follow-up duration between HPV⁺ arm and HPV⁻ arm (Student's *t*-test *p*-value = 0.25). Cohort participants were followed for an average of 6 visits, with a Poisson 95% CI from 5.63 to 6.43. The number of visits per participant ranges from 3 to 15, with an interquartile range (IQR) from 5 to 8. Women in the HPV⁻ participated in less visits than women in the HPV⁺ arm (Mann-Whitney U test *p*-value <0.001).

Participants were on average aged 21 at inclusion (median with an IQR of 20 to 23) and half of them (66/132) were vaccinated against HPV (56 with Gardasil, 6 with Cervarix, and 4 not specified). Additional characteristics can be found in Table S1.

For 73% of the women included in the analysis (97/132), we detected at least one episode (prevalent or incident) during their follow-up, and 45% (59/132) experienced codetections (i.e. the simultaneous detections of more than one genotype, whether prevalent or incident). About 61% (81/132) of the participants were already positive for at least one genotype at inclusion, and half of them were coinfecting at inclusion (39/132). In total, we detected 313 distinct episodes. Table 1 recapitulates the distribution of prevalent and incident episodes and Table 2 lists the occurrences of each of the 25 genotypes identified by our genotyping method. The three most frequently detected types in descending order were HPV51, HPV53, and HPV66. Half of the participants (64/132) exhibited 1 or less HPV genotype during follow-up, while the 68 remaining displayed various numbers of episodes, the largest being 11 distinct episodes for one participant (Figure S1). Overall, an average of 2.4 (Poisson 95% CI of 2.12 to 2.65) episodes per woman were detected during each follow-up, which yielded an average of 2.72 HPV episodes detected per person-year. We detected significantly fewer episodes in vaccinated women, with several genotypes being underrepresented in this group such as HPV16, HPV31, and HPV51. These discrepan-

cies between vaccinated and unvaccinated participants for the 25 genotypes are summarised in Table 2. There were no significant differences in the number of visits and duration of follow-up between vaccinated and unvaccinated participants.

Viral and socioeconomic factors’ association with HPV detection risk

The full results of the Cox analyses are shown in Figure 2. Briefly, we found that vaccinated participants were less likely to display new incident detection compared to unvaccinated participants, hazard ratio of 0.64 [0.43 to 0.96] (95%-CI). Similarly for first incident detection, vaccine-targeted genotypes (HPV6, HPV11, HPV16, HPV18) were less frequently detected compared to the other genotypes (Figure 2, log-rank p -value 0.007). Conversely, women reporting more than 3 LTSP had an increased probability of experiencing new detection compared to women reporting 1 or 2 LTSP (our baseline group). Note that although the magnitude of the effect is similar, the hazard ratio for the group reporting more than 10 LTSP is not significant, most likely due to the smaller population size.

When stratifying the rate of first incident detection by oncogenic risk, we found that HR-HPV were more likely to be detected than LR-HPV over the whole follow-up (log-rank p -value < 0.001).

Table 1: **Composition of the HPV-positive episodes detected.** Singleton refers to episodes remaining positive for a single visit only and FU to follow-up. N is the number of episodes.

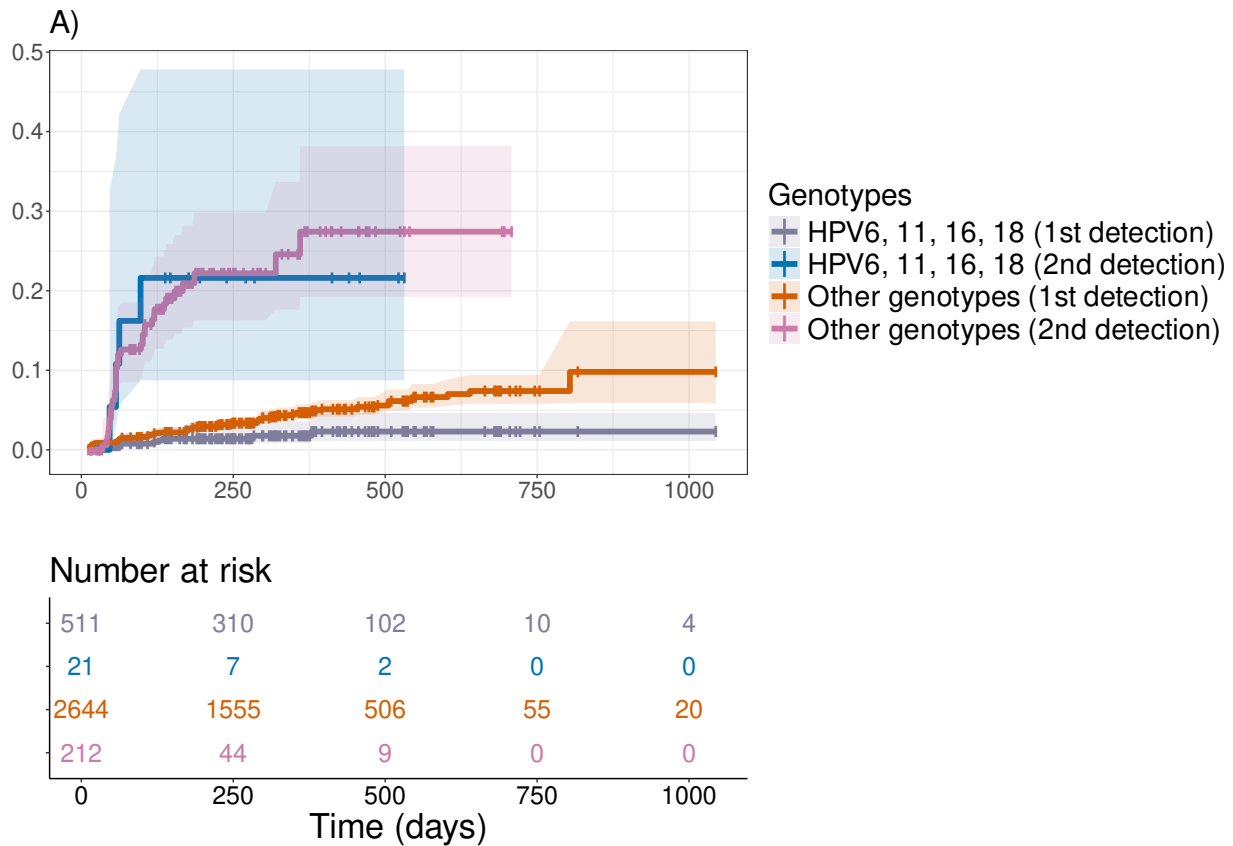
Episodes	Additional detail	N	%
Prevalent	all	145	46.33
	incl. positive over the entire FU	16	11.03
Incident	all	168	53.67
	incl. first occurrence	122	72.62
	incl. second occurrence	41	24.40
	incl. third occurrence	5	2.98
Singletons	all	176	56.23
	incl. censored observations	77	43.75
<i>All detected episodes</i>	all	313	100

Loss of HPV detection

Among the 145 prevalent episodes detected, 16 (11.03%) were positive for a genotype at inclusion and still positive for that same genotype when leaving the cohort, without any negative visit in between. These 16 episodes were detected in 13 women, who were followed for a shorter duration compared to the rest of the cohort (difference in the mean between the two groups of 130 days with a 95%CI from 25 to 235). Still, the majority of the episodes were positive for only one visit (176/313, i.e. 56.2%), even though a large proportion of these were censored observations (77/176, i.e. 43.8%). These results are summarised

Table 2: **Number of episodes detected in vaccinated and unvaccinated participants stratified by genotypes.** All incident and prevalent HPV-positive episodes were included in the table. The difference between the two groups was evaluated using a chi-square test. We display both raw and adjusted p -values and use the Benjamini-Hochberg procedure to correct for false discovery rates in multiple comparisons setting [29].

HPV genotype	Vaccinated	Unvaccinated	Total	p -value (raw)	p -value (corrected)
HPV6	1	4	5	0.180	0.479
HPV11	0	0	0	–	–
HPV16	3	11	14	0.033	0.260
HPV18	5	6	11	0.763	0.872
HPV31	4	16	20	0.007	0.087
HPV33	0	1	1	0.317	0.586
HPV34	1	1	2	1	1
HPV35	4	11	15	0.071	0.424
HPV39	8	10	18	0.637	0.765
HPV40	1	1	2	1	1
HPV42	1	4	5	0.180	0.431
HPV43	1	2	3	0.564	0.796
HPV44	7	3	10	0.206	0.449
HPV45	0	3	3	0.083	0.400
HPV51	10	28	38	0.004	0.084
HPV52	12	15	27	0.564	0.751
HPV53	14	19	33	0.384	0.658
HPV54	10	9	19	0.819	0.893
HPV56	8	11	19	0.491	0.737
HPV58	2	5	7	0.257	0.514
HPV59	7	2	9	0.096	0.287
HPV66	10	19	29	0.095	0.325
HPV70	2	1	3	0.564	0.712
HPV74	0	3	3	0.083	0.333
non-typable	10	7	17	0.467	0.747
All types	121	192	313	< 0.001	–



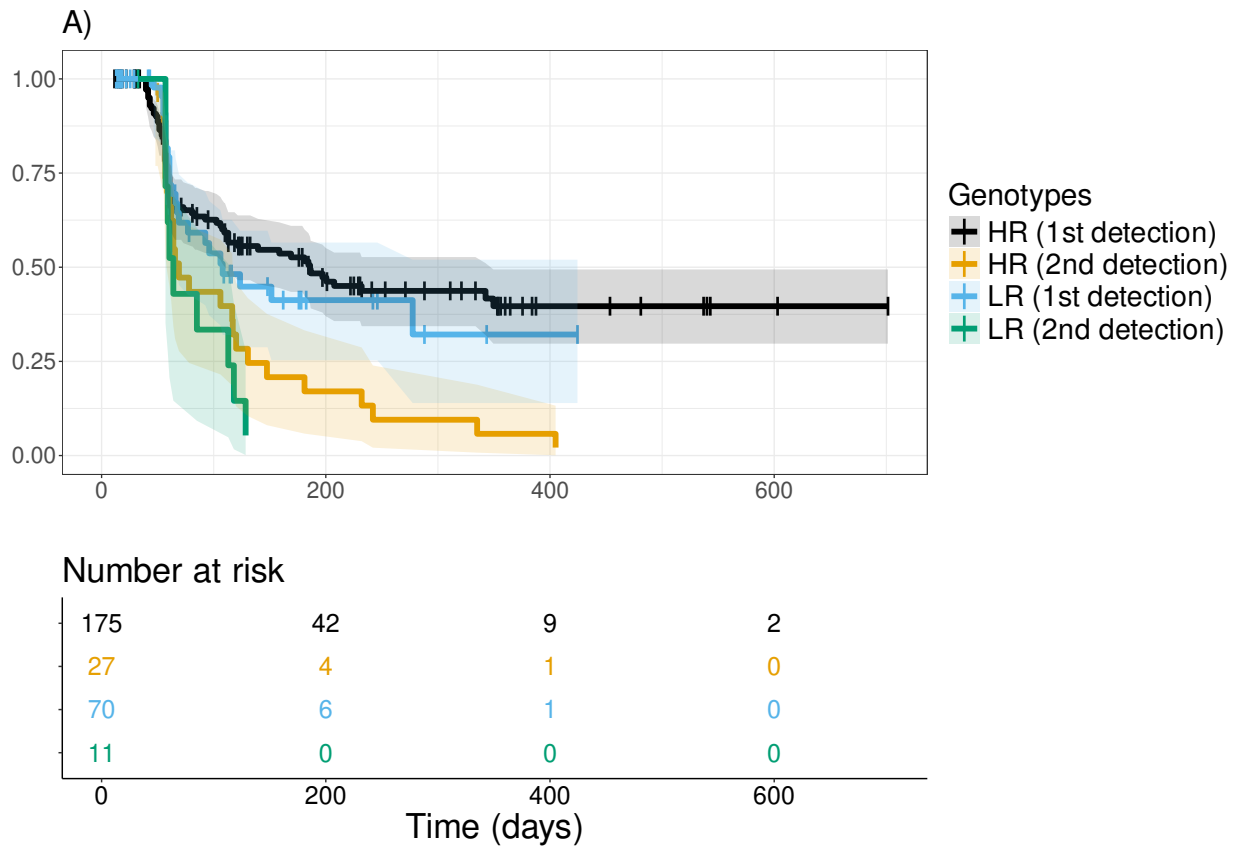
B)

	HR [CI 95%]
HPV vaccinated [vs unvaccinated]	0.63 [0.42, 0.96]
Heterosexual [vs bi-/homosexual]	0.86 [0.49, 1.51]
Caucasian origin [vs other ethnies]	0.88 [0.57, 1.34]
Financial difficulties [vs none]	0.95 [0.66, 1.36]
Use male condoms [vs none]	0.99 [0.88, 1.11]
Use oral contraceptive pills [vs none]	1.00 [0.85, 1.18]
More than 10 lifetime partners [ref: 1-2]	2.15 [0.93, 4.93]
3-10 lifetime partners [ref: 1-2]	2.40 [1.07, 5.39]

Figure 2: **Effect of genotype and host factors on the times to HPV first and second detection.** A) Cumulative distribution function (1 - survival function) of the time to first and second incident HPV detection, stratified by genotype (targeted by the quadrivalent vaccine or not). Thick lines show the estimated cumulative distribution functions and shaded areas the 95% log-log confidence intervals. Vertical ticks on the thick lines indicate an event was censored at the corresponding time. The table indicates the number of events at risk of occurring at each time step B) Averaged hazard ratio for the best models selected by Cox regression with frailty at the patient level. Significant covariates are in red and hazard ratios greater than 1 indicate the covariate is associated with an increased risk of HPV detection. The reference level is indicated in the bracket for the qualitative variables (see the Methods for details).

in Table 1.

Our estimates indicate a median period of HPV detectability of 113 days (log-log 95% CI of 96 to



B)

Factor	HR [CI 95%]
Financial difficulties [vs none]	0.45 [0.20, 0.97]
Caucasian origin [vs other ethnies]	0.54 [0.31, 0.93]
Heterosexual [vs bi-/homosexual]	0.97 [0.73, 1.28]
BMI at baseline visit	0.99 [0.97, 1.02]
Years between first menarchy and baseline visit	1.00 [0.96, 1.04]
More than 10 lifetime partners [ref: 1-2]	1.01 [0.51, 2.02]
Use male condoms [vs none]	1.07 [0.77, 1.47]
HPV vaccinated [vs unvaccinated]	1.09 [0.76, 1.57]
Use oral contraceptive pills [vs none]	1.15 [0.75, 1.74]
3-10 lifetime partners [ref: 1-2]	1.22 [0.55, 2.70]

Figure 3: **Effect of genotype oncogenic risk and host factors on the estimated duration of HPV detection.** A) Survival functions stratified by the genotype oncogenic risk (HR: high-risk ; LR: low-risk) for the duration of HPV detection. See Figure 2's caption for details B) Averaged hazard ratios for the host factors. Significant factors are in red and a hazard ratio lower than one indicates that the trait is associated with a decreased rate of recovery (loss of HPV DNA detection), hence longer survival functions. The reference level is indicated in the bracket for the qualitative variables (see Methods for details).

136) and that around 25.0% (log-log 95% CI of 17.5 to 33.2) of the HPV episodes were still detectable after 405 days of follow-up. The longest detected episodes lasted for at least 701 days (right-censored observation).

HPVs were detected for a significantly longer duration in participants who reported experiencing financial difficulties in the last 12 months before inclusion compared to participants who did not, hazard ratio of 0.45 [0.21-0.97] (95%-CI). The detection duration was also longer in participants who identified themselves solely as 'Caucasians', hazard ratio of 0.54 [0.32 to 0.96] (95%-CI), compared to participants who identified themselves as non-Caucasian or mixed-origin (Figure 3). We found no significant differences in the time to HPV clearance between high-risk HPV and low-risk HPV.

Discussion

Understanding HPV infection detection patterns and the factors associated with its variations is essential to optimise cervical cancer screening, guide future fundamental research in virology, and design efficient vaccination campaigns to improve vaccination coverage [30]. We addressed this issue by analysing HPV detection patterns in a cohort of 132 young women from Montpellier (France) with frequent and regular follow-ups. The design of this cohort study allows us to focus on the role of socioeconomic and viral factors in the detection and acquisition of new HPV infections.

Consistent with observations from a dozen countries [31], we found that vaccinated participants were less at risk of displaying incident infections than unvaccinated participants. Our result also suggests that HPV-16/18 AS04-adjuvanted vaccine (Cervarix[®], GlaxoSmithKline Biologicals) and HPV-6/11/16/18 vaccine (Gardasil[®], Merck) may prevent the acquisition of non-vaccine genotypes such as HPV31 or HPV51 (Table 2), even though the dataset here investigated is underpowered to highlight it. This cross-protective effect of bivalent and quadrivalent vaccines has been described elsewhere [32], [33], notably in regards to their phylogenetic distance from HPV16 and HPV18 [34]. Earlier work on the two first visits of the PAPCLEAR cohort pinpointed the cross-protective effect of the vaccination against HPV51 [35]. The longitudinal analysis conducted here gave us additional statistical power to support these trends. The vaccination coverage in the PAPCLEAR cohort (50%) is a bit higher than the vaccination coverage in mainland France (43.6% for the 15-18yo in 2021 [36]), but we expect similar trends over the whole territory for young women.

Compared to previous studies [8], [18], we observed more incident detections and redetections after 1 year and 2 years of follow-up. These differences could stem from the cohort design. First, participants' age (21.2 years old on average at inclusion) was generally lower than in some earlier studies [18], [37]. In adults, sexual activity is negatively correlated with age [38]. Therefore, we expect our study population to be more exposed to HPVs. Second, an inclusion criterion was to report at least one new sexual partner in the 12 months, which is also expected to increase HPV incidence [39], [40]. Finally, the dense frequency

between visits in the PAPCLEAR design (every 8 weeks, compared to 6 months in most studies) could also contribute to this trend. The frequency of codetections was also high (47%), but consistent with previous studies [8].

The time between two HPV positive events was statistically shorter for HR-HPV types compared to LR-HPV types, which is also consistent with results from other cohorts [8]. In our setting, we could not determine if this difference is due to the higher prevalence in the population of HR genotypes, increased rates of re-emergence from subclinical HR infections, or a combination of both. We lacked statistical power to assess whether this difference in survival functions between HR genotypes and LR genotypes was consistent among the second and third incident detections of the same genotype in a participant. There was also a tendency for HR-HPV types to be detected for a longer period than LR-HPV types (Figure 3), which is consistent with earlier studies [41], but the trend was not statistically significant (log-rank p -value of 0.13).

We found shorter mean durations of HPV detection than in earlier studies [8], [18], [37], [41]. This can be partially explained by differences in sampling rates and our inclusion of all positive episodes. Indeed, episodes consisting of a single observation, or ‘singletons’, are sometimes excluded from analyses [37]. Participants who reported having experienced financial difficulties in the past 12 months prior to their inclusion displayed longer periods of HPV positivity compared to those who did not report it. We expect the burden of precariousness to be even stronger in France than what we estimated as reports emphasised the reduced vaccination coverage rate in poorer households [36].

Participants who identified themselves as ‘Caucasian’ also experienced longer periods of HPV detection. However, we lacked data to assess if this trend reflected human genomics or socio-demographic/behavioural differences between the two groups. Besides, our sample size is relatively limited (27 reporting mixed origin or non-Caucasian origin and 105 Caucasian origin) and monocentric. This limited reflection of the general population diversity could generate a selection bias.

To date, except for some specific populations, e.g. women who remain abstinent, it is difficult to assess

the origin of a new HPV detection [9]. Therefore, it is generally delicate to extrapolate the dynamics of HPV infections from cervical smears HPV detection. However, earlier studies conducted in settings with only new sexual partners concluded that true incident infection is generally the most likely explanation [39], [40]. This is even more likely in young adults, for whom we have little evidence of the re-emergence of subclinical infections, and who are the most exposed to HPV infections [10]. Following these earlier studies, in our setting, a new detection would most likely correspond to a novel acquisition of the detected genotype.

One potential issue has to do with the rate of false negative detection. Here, we assumed that two episodes separated only by one negative visit were two distinct statistical events. Thus, the occurrences of false negative detection would bias our analysis towards shorter and more frequent detected HPV episodes. We relied on the use of the LiPA25 assay for its great sensitivity [42], precisely in order to limit false negatives. Furthermore, the addition of a DEIA test in the case of a LiPa-negative sample further helped avoid these false negative detections.

Still, there is no clear consensus in the choice of the threshold to distinguish two episodes caused by the same genotype in one individual and other works also used a threshold of two negative visits instead of one [37]. To evaluate the bias associated with such an assumption, we analysed our data with the two approaches (see Supplementary Materials A6). The two analyses yielded similar results and our previous conclusions remain unaffected.

More generally, there is a need to develop new statistical and clinical methods to assess accurately the status of HPV infection. In particular, exploring less invasive devices, such as tests on urine samples, can help to maximise participant enrolment and follow-up density, to better characterise the role of subclinical infection in HPV life history [11].

Acknowledgements

The authors thank all the participants of the PAPCLEAR study and the clinical staff and nurses for their help.

Role of the funding sources

The European Research Council and the Ligue Contre le Cancer had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

References

- [1] I. G. Bravo and M. Félez-Sánchez, “Papillomaviruses: Viral evolution, cancer and evolutionary medicine,” *Evol Med Public Health*, vol. 2015, no. 1, pp. 32–51, 2015. DOI: [10.1093/emph/eov003](https://doi.org/10.1093/emph/eov003).
- [2] H. Sung, J. Ferlay, R. L. Siegel, *et al.*, “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021, ISSN: 1542-4863. DOI: [10.3322/caac.21660](https://doi.org/10.3322/caac.21660).
- [3] H. W. Chesson, E. F. Dunne, S. Hariri, and L. E. Markowitz, “The estimated lifetime probability of acquiring human papillomavirus in the United States,” *Sexually Transmitted Diseases*, vol. 41, no. 11, pp. 660–664, 2014, ISSN: 1537-4521. DOI: [10.1097/OLQ.000000000000193](https://doi.org/10.1097/OLQ.000000000000193).
- [4] C. F. Houlihan, K. Baisley, I. G. Bravo, *et al.*, “Rapid acquisition of HPV around the time of sexual debut in adolescent girls in Tanzania,” *Int J Epidemiol*, vol. 45, no. 3, pp. 762–773, 2016. DOI: [10.1093/ije/dyv367](https://doi.org/10.1093/ije/dyv367).
- [5] L. Bruni, M. Diaz, M. Castellsagué, E. Ferrer, F. X. Bosch, and S. de Sanjosé, “Cervical Human Papillomavirus Prevalence in 5 Continents: Meta-Analysis of 1 Million Women with Normal Cyto-

- logical Findings,” *The Journal of Infectious Diseases*, vol. 202, no. 12, pp. 1789–1799, Dec. 2010, ISSN: 0022-1899. DOI: [10.1086/657321](https://doi.org/10.1086/657321). [Online]. Available: <https://doi.org/10.1086/657321>.
- [6] X. Castellsagué, T. Drudis, M. P. Cañadas, *et al.*, “Human Papillomavirus (HPV) infection in pregnant women and mother-to-child transmission of genital HPV genotypes: A prospective study in Spain,” *BMC infectious diseases*, vol. 9, p. 74, May 2009, ISSN: 1471-2334. DOI: [10.1186/1471-2334-9-74](https://doi.org/10.1186/1471-2334-9-74).
- [7] R. P. Insinga, E. J. Dasbach, E. H. Elbasha, K.-L. Liaw, and E. Barr, “Incidence and duration of cervical human papillomavirus 6, 11, 16, and 18 infections in young women: An evaluation from multiple analytic perspectives,” *Cancer Epidemiol Biomarkers Prev*, vol. 16, no. 4, pp. 709–15, 2007. DOI: [10.1158/1055-9965.EPI-06-0846](https://doi.org/10.1158/1055-9965.EPI-06-0846).
- [8] A. V. Ramanakumar, P. Naud, C. M. Roteli-Martins, *et al.*, “Incidence and duration of type-specific human papillomavirus infection in high-risk HPV-naïve women: Results from the control arm of a phase II HPV-16/18 vaccine trial,” *BMJ Open*, vol. 6, no. 8, e011371, 2016. DOI: [10.1136/bmjopen-2016-011371](https://doi.org/10.1136/bmjopen-2016-011371).
- [9] P. Gravitt and R. Winer, “Natural History of HPV Infection across the Lifespan: Role of Viral Latency,” *Viruses*, vol. 9, no. 10, p. 267, Sep. 2017, ISSN: 1999-4915. DOI: [10.3390/v9100267](https://doi.org/10.3390/v9100267). [Online]. Available: <http://www.mdpi.com/1999-4915/9/10/267>.
- [10] M. Schiffman, N. Wentzensen, S. Wacholder, W. Kinney, J. C. Gage, and P. E. Castle, “Human Papillomavirus Testing in the Prevention of Cervical Cancer,” *J Natl Cancer Inst*, vol. 103, no. 5, pp. 368–383, Mar. 2011, Publisher: Oxford Academic, ISSN: 0027-8874. DOI: [10.1093/jnci/djq562](https://doi.org/10.1093/jnci/djq562). [Online]. Available: <https://academic.oup.com/jnci/article/103/5/368/905734>.
- [11] J. Doorbar, “The human Papillomavirus twilight zone – Latency, immune control and subclinical infection,” *Tumour Virus Research*, vol. 16, p. 200–268, Dec. 2023, ISSN: 2666-6790. DOI: [10.1016/](https://doi.org/10.1016/)

[j.tvr.2023.200268](https://www.sciencedirect.com/science/article/pii/S2666679023000150). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666679023000150> (visited on 09/09/2023).

- [12] C. L. Murall, M. Rahmoun, C. Selinger, *et al.*, “Natural history, dynamics, and ecology of human papillomaviruses in genital infections of young women: Protocol of the PAPCLEAR cohort study,” *BMJ Open*, vol. 9, no. 6, Jun. 2019. DOI: [10.1136/bmjopen-2018-025129](https://doi.org/10.1136/bmjopen-2018-025129). [Online]. Available: <http://bmjopen.bmj.com/content/9/6/e025129.abstract>.
- [13] B. Kleter, L.-J. van Doorn, J. ter Schegget, *et al.*, “Novel Short-Fragment PCR Assay for Highly Sensitive Broad-Spectrum Detection of Anogenital Human Papillomaviruses,” *Am J Pathol*, vol. 153, no. 6, pp. 1731–1739, 1998, ISSN: 0002-9440. DOI: [10.1016/S0002-9440\(10\)65688-X](https://doi.org/10.1016/S0002-9440(10)65688-X).
- [14] D. T. Geraets, L. Struijk, B. Kleter, *et al.*, “The original SPF10 LiPA25 algorithm is more sensitive and suitable for epidemiologic HPV research than the SPF10 INNO-LiPA Extra,” *J Virol Meth*, vol. 215-216, pp. 22–29, 2015, ISSN: 0166-0934. DOI: [10.1016/j.jviromet.2015.01.001](https://doi.org/10.1016/j.jviromet.2015.01.001).
- [15] IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, *Biological Agents* (IARC monographs on the evaluation of carcinogenic risks to humans). Lyon, France: International Agency for Research on Cancer, 2012, vol. 100B, ISBN: 978-92-832-1319-2 978-92-832-0134-2.
- [16] F. Coutlée, P. Gravitt, J. Kornegay, *et al.*, “Use of PGMV Primers in L1 Consensus PCR Improves Detection of Human Papillomavirus DNA in Genital Samples,” *Journal of Clinical Microbiology*, vol. 40, no. 3, pp. 902–907, Mar. 2002. DOI: [10.1128/JCM.40.3.902-907.2002](https://doi.org/10.1128/JCM.40.3.902-907.2002).
- [17] J. Dillner, M. Arbyn, E. Unger, and L. Dillner, “Monitoring of human papillomavirus vaccination,” *Clin Exp Immunol*, vol. 163, no. 1, pp. 17–25, 2011. DOI: [10.1111/j.1365-2249.2010.04268.x](https://doi.org/10.1111/j.1365-2249.2010.04268.x).
- [18] T. Malagón, H. Trottier, M. El-Zein, L. Villa, and E. Franco, “Human papillomavirus intermittence and risk factors associated with first detections and redetections in the Ludwig-McGill cohort study of adult women,” *The Journal of infectious diseases*, Feb. 2023, Publisher: J Infect Dis, ISSN:

- 1537-6613. DOI: [10.1093/infdis/jiad043](https://pubmed.ncbi.nlm.nih.gov/36790831/). [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/36790831/>.
- [19] W. Nelson, “Hazard Plotting for Incomplete Failure Data,” *Journal of Quality Technology*, vol. 1, no. 1, pp. 27–52, 1969. DOI: [10.1080/00224065.1969.11980344](https://doi.org/10.1080/00224065.1969.11980344).
- [20] W. Nelson, “Theory and Applications of Hazard Plotting for Censored Failure Data,” *Technometrics*, vol. 14, no. 4, pp. 945–966, 1972. DOI: [10.1080/00401706.1972.10488991](https://doi.org/10.1080/00401706.1972.10488991).
- [21] O. Aalen, “Nonparametric Inference for a Family of Counting Processes,” *The Annals of Statistics*, vol. 6, no. 4, pp. 701–726, 1978. DOI: [10.1214/aos/1176344247](https://doi.org/10.1214/aos/1176344247).
- [22] N. Mantel, “Evaluation of survival data and two new rank order statistics arising in its consideration,” eng, *Cancer Chemotherapy Reports*, vol. 50, no. 3, pp. 163–170, Mar. 1966, ISSN: 0069-0112.
- [23] R. Peto and J. Peto, “Asymptotically Efficient Rank Invariant Test Procedures,” *Journal of the Royal Statistical Society. Series A (General)*, vol. 135, no. 2, pp. 185–207, 1972. DOI: [10.2307/2344317](https://doi.org/10.2307/2344317).
- [24] D. R. Cox, “Regression Models and Life-Tables,” in *Breakthroughs in Statistics: Methodology and Distribution*, ser. Springer Series in Statistics, S. Kotz and N. L. Johnson, Eds., New York, NY: Springer, 1992, pp. 527–541. DOI: [10.1007/978-1-4612-4380-9_37](https://doi.org/10.1007/978-1-4612-4380-9_37).
- [25] A. Wienke, *Frailty Models in Survival Analysis*. New York: Chapman and Hall/CRC, Sep. 2010, ISBN: 978-0-429-13960-4. DOI: [10.1201/9781420073911](https://doi.org/10.1201/9781420073911).
- [26] K. P. Burnham and D. R. Anderson, “Multimodel Inference: Understanding AIC and BIC in Model Selection,” *Sociological Methods & Research*, vol. 33, no. 2, pp. 261–304, 2004. DOI: [10.1177/0049124104268644](https://doi.org/10.1177/0049124104268644).
- [27] M. R. E. Symonds and A. Moussalli, “A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike’s information criterion,” *Behavioral Ecology and Sociobiology*, vol. 65, no. 1, pp. 13–21, 2011. DOI: [10.1007/s00265-010-1037-6](https://doi.org/10.1007/s00265-010-1037-6).

- [28] C. M. Hurvich and C.-L. Tsai, “A Corrected Akaike Information Criterion for Vector Autoregressive Model Selection,” *Journal of Time Series Analysis*, vol. 14, no. 3, pp. 271–279, 1993. DOI: [10.1111/j.1467-9892.1993.tb00144.x](https://doi.org/10.1111/j.1467-9892.1993.tb00144.x).
- [29] Y. Benjamini and Y. Hochberg, “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing,” en, *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995, ISSN: 2517-6161. DOI: [10.1111/j.2517-6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x). [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02031.x> (visited on 08/06/2024).
- [30] J. E. Graham and A. Mishra, “Global challenges of implementing human papillomavirus vaccines,” *International Journal for Equity in Health*, vol. 10, no. 1, p. 27, Jun. 2011, ISSN: 1475-9276. DOI: [10.1186/1475-9276-10-27](https://doi.org/10.1186/1475-9276-10-27). [Online]. Available: <https://doi.org/10.1186/1475-9276-10-27>.
- [31] J. M. L. Brotherton, “Impact of HPV vaccination: Achievements and future challenges,” eng, *Papillomavirus Research (Amsterdam, Netherlands)*, vol. 7, pp. 138–140, Jun. 2019, ISSN: 2405-8521. DOI: [10.1016/j.pvr.2019.04.004](https://doi.org/10.1016/j.pvr.2019.04.004).
- [32] T. Malagón, M. Drolet, M.-C. Boily, *et al.*, “Cross-protective efficacy of two human papillomavirus vaccines: A systematic review and meta-analysis,” *The Lancet Infectious Diseases*, vol. 12, no. 10, pp. 781–789, Oct. 2012, ISSN: 1473-3099. DOI: [10.1016/S1473-3099\(12\)70187-1](https://doi.org/10.1016/S1473-3099(12)70187-1). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1473309912701871>.
- [33] C. M. Wheeler, X. Castellsagué, S. M. Garland, *et al.*, “Cross-protective efficacy of HPV-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by non-vaccine oncogenic HPV types: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial,” eng, *The Lancet. Oncology*, vol. 13, no. 1, pp. 100–110, Jan. 2012, ISSN: 1474-5488. DOI: [10.1016/S1470-2045\(11\)70287-X](https://doi.org/10.1016/S1470-2045(11)70287-X).

- [34] J. A. Bogaards, P. van der Weele, P. J. Woestenbergh, B. H. B. van Benthem, and A. J. King, “Bivalent Human Papillomavirus (HPV) Vaccine Effectiveness Correlates With Phylogenetic Distance From HPV Vaccine Types 16 and 18,” *J Infect Dis*, vol. 220, no. 7, pp. 1141–1146, 2019. DOI: [10.1093/infdis/jiz280](https://doi.org/10.1093/infdis/jiz280).
- [35] C. L. Murall, B. Reyn e, C. Selinger, *et al.*, “HPV cervical infections and serological status in vaccinated and unvaccinated women,” *Vaccine*, vol. 38, no. 51, pp. 8167–8174, 2020. DOI: [10.1016/j.vaccine.2020.10.078](https://doi.org/10.1016/j.vaccine.2020.10.078).
- [36] R. Hanguhard, A. Gautier, N. Soullier, A. Barret, I. Parent du Chatelet, and S. Vaux, “Couverture vaccinale contre les infections   papillomavirus humain des filles  g es de 15   18 ans et d terminants de vaccination,” EN, *Bull  pid miol Hebd.*, 2021. [Online]. Available: http://beh.santepubliquefrance.fr/beh/2022/24-25/2022_24-25_3.html.
- [37] P. Paul, A. Hammer, A. F. Rositch, *et al.*, “Rates of New Human Papillomavirus Detection and Loss of Detection in Middle-aged Women by Recent and Past Sexual Behavior,” *The Journal of Infectious Diseases*, vol. 223, no. 8, pp. 1423–1432, Apr. 2021, ISSN: 0022-1899. DOI: [10.1093/infdis/jiaa557](https://doi.org/10.1093/infdis/jiaa557). [Online]. Available: <https://doi.org/10.1093/infdis/jiaa557>.
- [38] D. Herbenick, M. Reece, V. Schick, S. A. Sanders, B. Dodge, and J. D. Fortenberry, “Sexual behavior in the United States: Results from a national probability sample of men and women ages 14-94,” eng, *The Journal of Sexual Medicine*, vol. 7 Suppl 5, pp. 255–265, Oct. 2010, ISSN: 1743-6109. DOI: [10.1111/j.1743-6109.2010.02012.x](https://doi.org/10.1111/j.1743-6109.2010.02012.x).
- [39] H. Trottier, S. Ferreira, P. Thomann, *et al.*, “Human papillomavirus infection and reinfection in adult women: The role of sexual activity and natural immunity,” eng, *Cancer Research*, vol. 70, no. 21, pp. 8569–8577, Nov. 2010, ISSN: 1538-7445. DOI: [10.1158/0008-5472.CAN-10-0621](https://doi.org/10.1158/0008-5472.CAN-10-0621).

- [40] A.-B. Moscicki, Y. Ma, S. Farhat, *et al.*, “Redetection of cervical human papillomavirus type 16 (HPV16) in women with a history of HPV16,” eng, *The Journal of Infectious Diseases*, vol. 208, no. 3, pp. 403–412, Aug. 2013, ISSN: 1537-6613. DOI: [10.1093/infdis/jit175](https://doi.org/10.1093/infdis/jit175).
- [41] M. T. Goodman, Y. B. Shvetsov, K. McDuffie, *et al.*, “Prevalence, Acquisition, and Clearance of Cervical Human Papillomavirus Infection among Women with Normal Cytology: Hawaii Human Papillomavirus Cohort Study,” *Cancer Research*, vol. 68, no. 21, pp. 8813–8824, 2008. DOI: [10.1158/0008-5472.CAN-08-1380](https://doi.org/10.1158/0008-5472.CAN-08-1380).
- [42] D. T. Geraets, L. Struijk, B. Kleter, *et al.*, “The original SPF10 LiPA25 algorithm is more sensitive and suitable for epidemiologic HPV research than the SPF10 INNO-LiPA Extra,” eng, *Journal of Virological Methods*, vol. 215-216, pp. 22–29, Apr. 2015, ISSN: 1879-0984. DOI: [10.1016/j.jviromet.2015.01.001](https://doi.org/10.1016/j.jviromet.2015.01.001).

Supplementary Methods

A1 Ethics

The PAPCLEAR study was promoted by the Centre Hospitalier Universitaire de Montpellier and approved by the *Comité de Protection des Personnes (CPP) Sud Méditerranée I* on 11 May 2016 (CPP number 16 42, reference number ID RCB 2016A00712-49); by the *Comité Consultatif sur le Traitement de l'Information en matière de Recherche dans le domaine de la Santé* on 12 July 2016 (reference number 16.504); and by the *Commission Nationale Informatique et Libertés* on 16 December 2016 (reference number MMS/ABD/AR1612278, decision number DR-2016488). This study was authorised by the *Agence Nationale de Sécurité du Médicament et des Produits de Santé* on 20 July 2016 (reference 20160072000007). The ClinicalTrials.gov identifier is NCT02946346. All participants provided written informed consent.

A2 Defining events and duration

Results from the DEIA and LiPA25 assays yielded dated binary vectors. For each infectious event, we only know the intervals during which the infection started and ended, which means the data is ‘doubly interval censored’ and usually cumbersome to analyse [1]. To simplify the problem, we computed duration using the conventional midpoint methodology. For this, we defined the start of an infection as the midpoint between the last negative test before and the first positive test of the infection. Likewise, we defined the end of an infection as the midpoint between the last positive test of the infection and the first negative test after the infection). For incomplete data, we assumed the start to be at inclusion for left-censored observation and the end to be at the last visit for right-censored observation. The bias associated with this simpler method is expected to be limited since our sampling scheme is regular and short-spaced [2].

To study the time to HPV infection clearance, we defined an ‘event’ (or ‘episode’) as a series of at least one positive LiPA25 detection for a given HPV type and a given participant. During the follow-up,

we often detected several events per participant (sometimes even by the same genotype). We assumed that two consecutive episodes were independent even if only separated by one negative visit.

We computed the length of HPV detection as the duration between the midpoint at the start of an infection and the midpoint at the end of the infection. If one or both of the endpoints were censored, we assumed the duration to be right-censored. We assumed the events to be independent [3] and, therefore, defined the duration between episodes to be independent events. When analysing the time between positive episodes, we included all events and computed it as the duration between the midpoint at the start of the expired episode (or inclusion if it is the first incident event) and the midpoint at the start of the new episode. There were in general, fewer redetections than expected because some participants were still positive for a genotype at the end of their follow-up. If the genotype was not detected during follow-up, we used a right-censored observation with a duration equal to the follow-up duration of the participant. In both cases, the survival functions (hence the cumulative distribution functions) were computed using the Nelson-Aalen estimator of the cumulative hazard rate function [4]–[6].

A3 Statistical analyses

In our study, the unit of observation is the HPV genotype at the patient level. Therefore, the codetection of multiple genotypes in the same sample can induce some correlations between the observations. To account for correlations between observations of the same cluster (i.e. a participant), we added shared frailty effects at the patient level in the Cox regression [7] and used a likelihood ratio test with one degree of freedom between the two models (with and without the frailty effect) to identify the most relevant one. For all Cox regression analyses, we checked the validity of the proportional hazards (PH) assumption using Schoenfeld’s residuals [8]. We first generated the ‘maximum’ model, which contained all the variables chosen for the Cox regression, and then performed our model selection by considering all possible combinations from the maximum model and evaluating their respective AICc. We kept the models with an AICc smaller than the minimum AICc+2, following standard practice [9]. We then averaged the

coefficients of the remaining models using a full averaging procedure to avoid artificial departure from 0. This was necessary because we averaged on all the selected models, not just on the ones with the variable whose coefficient was computed [10]. Finally, we computed the hazard ratio by taking the exponential of these averaged coefficients.

A4 Characteristics of PAPCLEAR participants

Table S1: **Characteristics of the participants included in the analysis.** Except for the type of HPV vaccine, which was ignored, the first level displayed for all categorical variables was used as the reference level in the Cox analyses. Missing observations were removed from the analysis.

Quantitative covariates	Median [IQR]	women (<i>N</i>)	women (%)
Age at inclusion	21 [20 - 24]	132	100
Age at first menstruation	13 [12 - 14]	131	99.24
	Missing	1	0.76
Years between 1 st menstruation and inclusion	8 [7 - 10]	131	99.24
	Missing	1	0.76
Age of first sexual intercourse	16.5 [15 - 17]	132	100
Years between 1 st intercourse and inclusion	5 [3 - 7]	132	100

Categorical covariates	Category	women (<i>N</i>)	women (%)
HPV Vaccination	Unvaccinated	66	50.00
	Vaccinated	66	50.00
Vaccine used	Cervarix	6	9.09
	Gardasil4	56	84.18
	Missing	4	6.06
Ethnicity	Mixed or other	27	20.45
	Caucasian	105	79.55
Sexual preferences	Bi-/Homosexual	12	9.09
	Heterosexual	120	90.90
Smoking status	Never	61	46.21
	Past	19	14.39
	Current	52	39.39
Oral hormonal contraception [†]	Not using	64	48.48
	User	68	51.51
Condom use by male partner	Not using	56	42.42
	User	76	57.57
Number of lifetime sexual partners	1-2	20	15.15
	3-10	63	47.72
	11+	48	36.36
	Missing	1	0.76
Financial difficulties*	None	117	88.64
	Experienced in the last 12mo	15	11.36

† emergency pills not included

* approximated by having forgone medical care because of financial reasons in the 12 months prior to the baseline visit

A5 Distribution of number of episodes detected for each participant

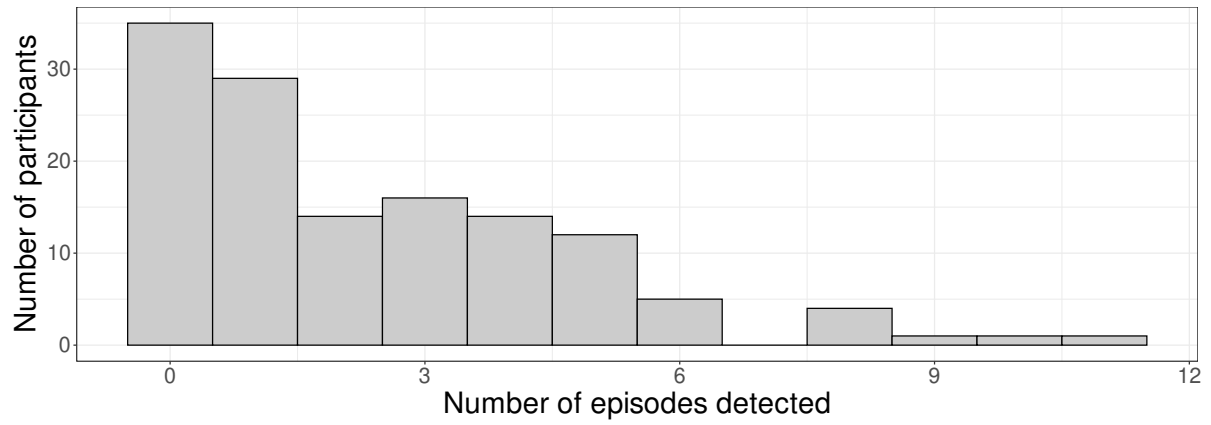
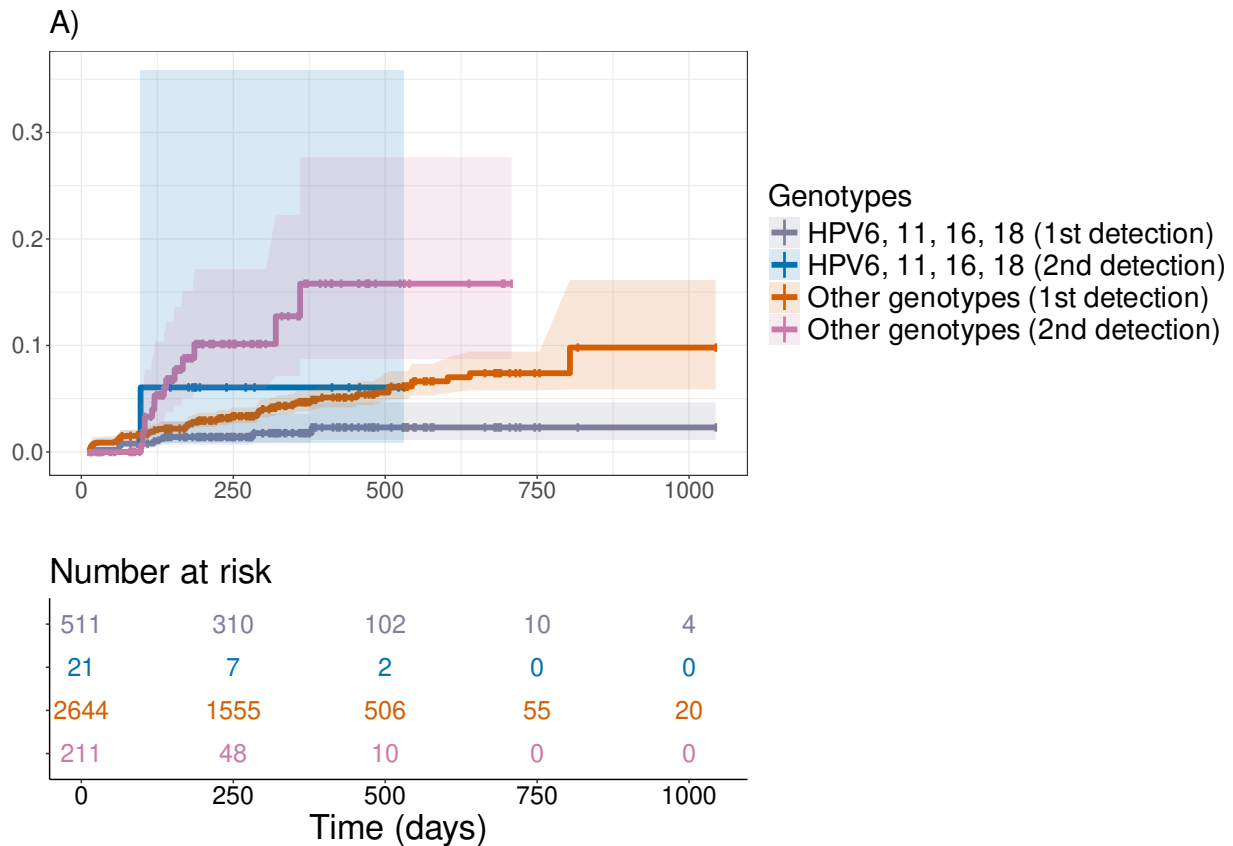


Figure S1: **Distribution of the number of episodes detected per participant.**

A6 Merging intermittent patterns

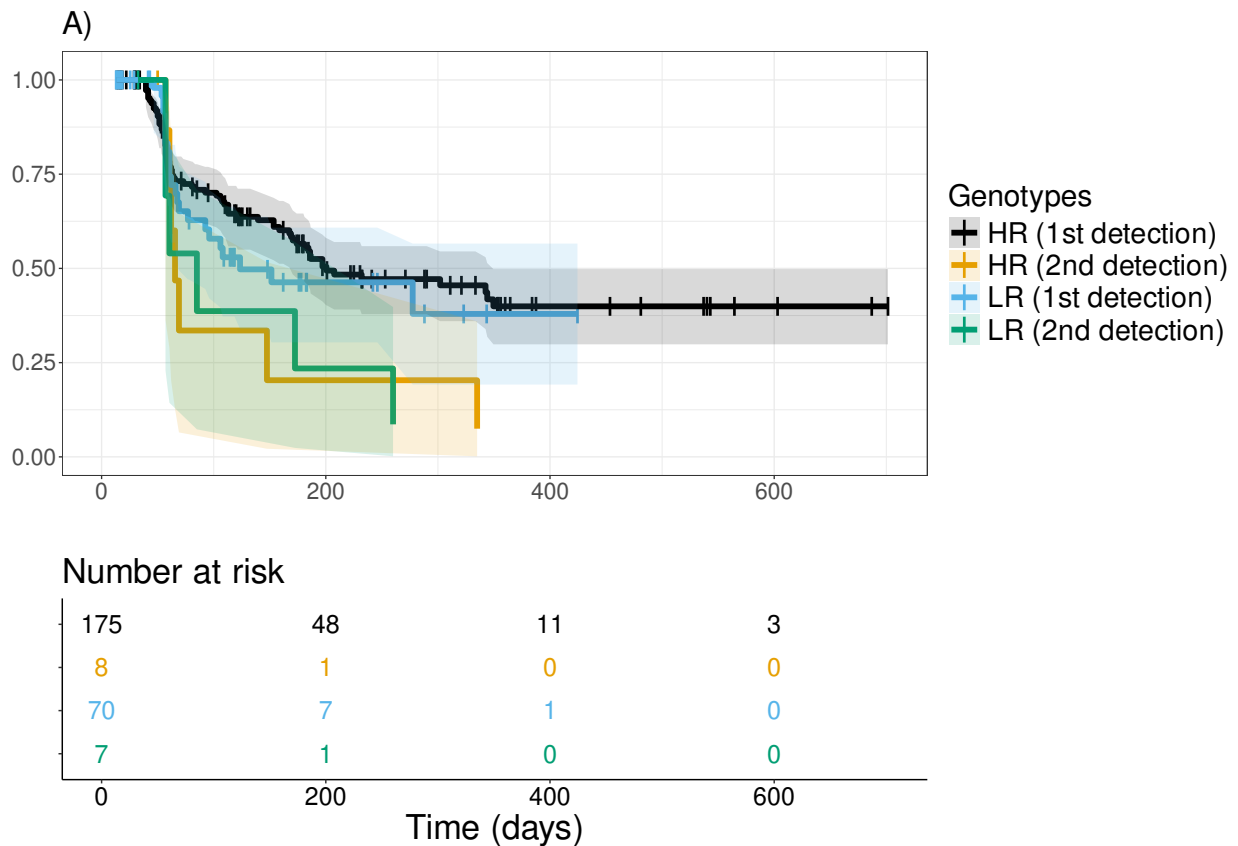
Following previous notations, intermittent patterns correspond to successive positive HPV detection episodes separated only by one negative visit [11]. There is no clear consensus on the definition of an HPV detection episode. In the main document, we consider two episodes as distinct if separated by at least one HPV genotype-negative visit. Others have used a threshold of two negative visits to distinguish between distinct consecutive episodes [12]. Merging intermittent patterns modifies the data used for analysis by diminishing the number of events and making them last longer on average. In total, we detected 31 intermittent patterns. Merging these patterns decreased the number of positive detectable events by the same amount. However, it did not qualitatively affect the results of the Cox regression but changed degrees of significance (see Figure S2 for the time between episodes and Figure S3 for the duration of detection). Notably, the contribution of financial difficulties to the increased risk of HPV detection is no longer significant.



B)

	HR [CI 95%]
HPV vaccinated [vs unvaccinated]	0.59 [0.37, 0.93]
Caucasian origin [vs other ethnies]	0.64 [0.38, 1.09]
Financial difficulties [vs none]	0.82 [0.43, 1.57]
Heterosexual [vs bi-/homosexual]	0.91 [0.56, 1.49]
Use male condoms [vs none]	0.95 [0.73, 1.24]
Years between first intercourse and baseline visit	0.99 [0.96, 1.02]
Use oral contraceptive pills [vs none]	1.07 [0.76, 1.50]
More than 10 lifetime partners [ref: 1-2]	2.62 [0.99, 6.88]
3-10 lifetime partners [ref: 1-2]	2.99 [1.18, 7.60]

Figure S2: **Effect of genotype and host factors on the times to HPV first and second detection with merged intermittent patterns.** A) Cumulative distribution function (1 - survival function) of the time to first and second incident HPV detection, stratified by genotype (targeted by the quadrivalent vaccine or not). Thick lines show the estimated cumulative distribution functions and shaded areas the 95% log-log confidence intervals. Vertical ticks on the thick lines indicate an event was censored at the corresponding time. The table indicates the number of events at risk of occurring at each time step B) Averaged hazard ratio for the best models selected by Cox regression with frailty at the patient level. Significant covariates are in red and hazard ratios greater than 1 indicate the covariate is associated with an increased risk of HPV detection. The reference level is indicated in the bracket for the qualitative variables (see the Methods for details).



B)

Factor	HR [CI 95%]
Caucasian origin [vs other ethnies]	0.51 [0.32, 0.82]
Financial difficulties [vs none]	0.58 [0.26, 1.25]
Heterosexual [vs bi-/homosexual]	0.98 [0.81, 1.18]
Years between first intercourse and baseline visit	1.00 [0.97, 1.02]
Use male condoms [vs none]	1.00 [0.89, 1.14]
Years between first menarchy and baseline visit	1.01 [0.95, 1.07]
More than 10 lifetime partners [ref: 1-2]	1.02 [0.45, 2.31]
Use oral contraceptive pills [vs none]	1.13 [0.76, 1.68]
HPV vaccinated [vs unvaccinated]	1.21 [0.80, 1.83]
3-10 lifetime partners [ref: 1-2]	1.29 [0.52, 3.18]

Figure S3: **Effect of genotype oncogenic risk and host factors on the estimated duration of HPV detection with merged intermittent detection patterns.** A) Survival functions stratified by the genotype oncogenic risk (HR: high-risk ; LR: low-risk) for the duration of HPV detection. See Figure S2's caption for details B) Averaged hazard ratios for the host factors. Significant factors are in red and a hazard ratio lower than one indicates that the trait is associated with a decreased rate of recovery (loss of HPV DNA detection), hence longer survival functions. The reference level is indicated in the bracket for the qualitative variables (see Methods for details).

A7 R packages

- SURVIVAL: non-parametric and parametric estimators of the survival function and Cox regression [13] ; version 3.5-3.
- MUMIN: model selection and model averaging [14] ; version 1.46.0.
- COXME: adding frailty effects to the hazard function as a centred Gaussian distribution [15] ; version 2.2-17.

References

- [1] K. Bogaerts, A. Komárek, and E. Lesaffre, *Survival Analysis with Interval-Censored Data*. Chapman and Hall/CRC, Nov. 2017, ISBN: 978-1-315-11694-5. DOI: [10.1201/9781315116945](https://doi.org/10.1201/9781315116945).
- [2] C. G. Law and R. Brookmeyer, “Effects of mid-point imputation on the analysis of doubly censored data,” *Statistics in Medicine*, vol. 11, no. 12, 1992, ISSN: 02776715. DOI: [10.1002/sim.4780111204](https://doi.org/10.1002/sim.4780111204).
- [3] J. Dillner, M. Arbyn, E. Unger, and L. Dillner, “Monitoring of human papillomavirus vaccination,” *Clin Exp Immunol*, vol. 163, no. 1, pp. 17–25, 2011. DOI: [10.1111/j.1365-2249.2010.04268.x](https://doi.org/10.1111/j.1365-2249.2010.04268.x).
- [4] W. Nelson, “Hazard Plotting for Incomplete Failure Data,” *Journal of Quality Technology*, vol. 1, no. 1, pp. 27–52, 1969. DOI: [10.1080/00224065.1969.11980344](https://doi.org/10.1080/00224065.1969.11980344).
- [5] W. Nelson, “Theory and Applications of Hazard Plotting for Censored Failure Data,” *Technometrics*, vol. 14, no. 4, pp. 945–966, 1972. DOI: [10.1080/00401706.1972.10488991](https://doi.org/10.1080/00401706.1972.10488991).
- [6] O. Aalen, “Nonparametric Inference for a Family of Counting Processes,” *The Annals of Statistics*, vol. 6, no. 4, pp. 701–726, 1978. DOI: [10.1214/aos/1176344247](https://doi.org/10.1214/aos/1176344247).
- [7] A. Wienke, *Frailty Models in Survival Analysis*. New York: Chapman and Hall/CRC, Sep. 2010, ISBN: 978-0-429-13960-4. DOI: [10.1201/9781420073911](https://doi.org/10.1201/9781420073911).
- [8] D. Schoenfeld, “Partial Residuals for The Proportional Hazards Regression Model,” *Biometrika*, vol. 69, no. 1, p. 239, 1982. DOI: [10.2307/2335876](https://doi.org/10.2307/2335876).

- [9] K. P. Burnham and D. R. Anderson, “Multimodel Inference: Understanding AIC and BIC in Model Selection,” *Sociological Methods & Research*, vol. 33, no. 2, pp. 261–304, 2004. DOI: [10.1177/0049124104268644](https://doi.org/10.1177/0049124104268644).
- [10] M. R. E. Symonds and A. Moussalli, “A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike’s information criterion,” *Behavioral Ecology and Sociobiology*, vol. 65, no. 1, pp. 13–21, 2011. DOI: [10.1007/s00265-010-1037-6](https://doi.org/10.1007/s00265-010-1037-6).
- [11] T. Malagón, H. Trottier, M. El-Zein, L. Villa, and E. Franco, “Human papillomavirus intermittence and risk factors associated with first detections and redetections in the Ludwig-McGill cohort study of adult women,” en, *The Journal of infectious diseases*, Feb. 2023, Publisher: J Infect Dis, ISSN: 1537-6613. DOI: [10.1093/infdis/jiad043](https://doi.org/10.1093/infdis/jiad043). [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/36790831/>.
- [12] P. Paul, A. Hammer, A. F. Rositch, *et al.*, “Rates of New Human Papillomavirus Detection and Loss of Detection in Middle-aged Women by Recent and Past Sexual Behavior,” *The Journal of Infectious Diseases*, vol. 223, no. 8, pp. 1423–1432, Apr. 2021, ISSN: 0022-1899. DOI: [10.1093/infdis/jiaa557](https://doi.org/10.1093/infdis/jiaa557). [Online]. Available: <https://doi.org/10.1093/infdis/jiaa557>.
- [13] T. M. Therneau, *A package for survival analysis in r*, R package version 3.4-0, 2022. [Online]. Available: <https://CRAN.R-project.org/package=survival>.
- [14] K. Bartoń, *Mumin: Multi-model inference*, R package version 1.46.0, 2022. [Online]. Available: <https://CRAN.R-project.org/package=Mumin>.
- [15] T. M. Therneau, *Coxme: Mixed effects cox models*, R package version 2.2-17, 2022. [Online]. Available: <https://CRAN.R-project.org/package=coxme>.