UDC 616-022.7:578.891.1(477)
DOI https://doi.org/10.32782/2226-2008-2026-3-5

A. I. Rybin *https://orcid.org/0000-0002-1145-6690*
V. M. Sokolov *https://orcid.org/0000-0003-4521-147X*
V. V. Osyk *https://orcid.org/0009-0009-2370-3677*
V. I. Zamyshlyak *https://orcid.org/0009-0001-1588-182X*
D. V. Sokolov *https://orcid.org/0000-0001-7360-5180*

CURRENT EPIDEMIOLOGICAL ASPECTS OF HUMAN PAPILLOMAVIRUS INFECTION IN UKRAINE

Odesa National Medical University, Odesa, Ukraine

UDC 616-022.7:578.891.1(477)

A. I. Rybin, V. M. Sokolov, V. V. Osyk, V. I. Zamyshlyak, D. V. Sokolov
CURRENT EPIDEMIOLOGICAL ASPECTS OF HUMAN PAPILLOMAVIRUS INFECTION IN UKRAINE
Odesa National Medical University, Odesa, Ukraine

Cervical cancer remains one of the leading causes of cancer morbidity and mortality among women worldwide, with persistent infection by high-risk human papillomavirus (HPV) types being the primary etiological factor. In Ukraine, cervical cancer consistently ranks among the top five oncological diseases in women, with over 3.000 new cases annually. The aim of this study was to assess the prevalence of HPV infection among women in different regions of Ukraine.

Material and Methods. The Quintip Recordati database data, containing 195 records, was analysed using frequency analysis methods. The distribution features of different HPV strains among the surveyed women living in other regions were determined. The frequency of HPV-positivity was analysed depending on the degree of urbanisation. The main ways of preventing HPV infection were determined.

Results. It was found that the number of positive tests was 38.5 % of the total sample size. When comparing the frequency of detection of highly oncogenic strains, it was found that they are more often registered in cities with a larger population.

Conclusions:

1. The prevalence of HPV in the examined women is twice the average European level.
2. The frequency of detection of highly oncogenic HPV strains is higher in urbanised regions.
3. Widespread implementation of HPV screening methods in Ukraine can significantly reduce the incidence of cervical cancer.

Keywords: human papillomavirus, epidemiology, women's health, prevention, screening.

УДК 616-022.7:578.891.1(477)

А. І. Рибін, В. М. Соколов, В. В. Осик, В. І. Замишляк, Д. В. Соколов
СУЧАСНІ АСПЕКТИ ЕПІДЕМІОЛОГІЇ ПАПІЛОМАВІРУСНОЇ ІНФЕКЦІЇ В УКРАЇНІ
Одеський національний медичний університет, Одеса, Україна

Мета дослідження – оцінити поширеність ВПЛ серед жінок у різних регіонах України. Проведено аналіз бази даних Quintip Recordati (195 записів) із застосуванням методів частотного аналізу та кластеризації. Встановлено, що частота виявлення ВПЛ становила 38,5 %, при цьому високоонкогенні типи частіше виявлялися у великих містах. Ідентифіковано чотири епідеміологічні кластери з різними рівнями активності ВПЛ. Висока поширеність інфекції в Україні пов'язана з низьким охопленням вакцинацією та недостатнім рівнем скринінгу. Розширення програм тестування та профілактики може суттєво знизити захворюваність на рак шийки матки.

Ключові слова: вірус папіломи людини, епідеміологія, жіноче здоров'я, профілактика, скринінг.

Introduction

Cervical cancer (CC) is one of the most common malignancies in women of reproductive age and ranks fourth among causes of cancer incidence and mortality worldwide. According to GLOBOCAN, in 2020, over 604.000 new cases and approximately 342.000 deaths were reported [1]. In Ukraine, CC consistently ranks among the five most common cancers in women, with over 3.000 new cases and more than 1.000 deaths annually [1].

Persistent infection with high-risk human papillomavirus (HPV) types, particularly 16 and 18, plays a key role in CC development by integrating into the genome of cervical epithelial cells and initiating molecular processes leading to neoplasia. However, not every HPV infection leads to cancer, highlighting the importance of immune response, comorbidities, and socio-behavioral factors [1]. The long natural history of cervical cell transformation provides a window for early intervention and prevention through vaccination, cytological screening, and HPV testing [2; 3].

HPV epidemiology shows differences between rural and urban populations, influenced by access to healthcare, socio-economic status, and cultural factors [3–8]. A Ukrainian study (2014–2016) demonstrated that HPV was detected in 34 % of women, with 65 % of infections attributed to high-risk genotypes, most frequently HPV-16; CIN III was predominantly associated with HPV-16(70 %) [9; 10].

© A. I. Rybin, V. M. Sokolov, V. V. Osyk et al., 2026

Стаття поширюється на умовах ліцензії

Thus, investigating the mechanisms of HPV-induced carcinogenesis and evaluating the effectiveness of prevention, screening, and treatment strategies for CC remain highly relevant for medicine, public health, and healthcare policy, taking into account rural–urban disparities [1–12].

The aim of this study was to assess the prevalence of HPV infection among women in different regions of Ukraine.

Materials and Methods

This study was designed as a retrospective, cross-sectional analysis, using routinely collected multi-center clinical data. The study adhered to the ethical principles of the Declaration of Helsinki (World Medical Association, 2013 revision) and was approved by the Bioethics Committee of Odesa National Medical University (Protocol No. 29, 12 April 2021). All records were de-identified at the point of database extraction by the data custodian.

The primary data source was the QVINTIP Recordati database [2], a structured registry containing HPV genotyping records from women across multiple regions of Ukraine. The analytical dataset comprised 195 records remaining after quality screening. Each record included: region of residence (with urbanisation classification), HPV genotype-level results (presence/absence and quantitative viral load where available), and date of testing (enabling temporal feature derivation).

Records were excluded if genotype results were technically invalid (internal control failure), the regional identifier was missing or ambiguous, or duplicate records from the same individual on the same date were identified (retaining only the most recent entry).

All HPV genotypes detected in the dataset were classified into oncogenic risk categories prior to analysis according to the IARC Monograph 100B classification [11] and subsequent updates: (I) Group 1 – confirmed carcinogenic: HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68; (II) Group 2A – probably carcinogenic: HPV 68 (reclassified in updated reviews); (III) Group 2B – possibly carcinogenic: HPV 26, 53, 66, 67, 70, 73, 85; (IV) low-risk (not classifiable as carcinogenic): HPV 6, 11, 40, 42, 43, 44, 54.

Genotype-level frequency analysis was conducted as the first analytical step to establish univariate prevalence estimates. For each HPV genotype and for the aggregate high-risk and low-risk genotype groups, the following statistics were computed: (I) absolute number of positive records; (II) prevalence proportion (% of total records); (III) stratified prevalence by urbanisation category (metropolitan, city, town, rural) and by region. Urbanisation classification followed the Ukrainian State Classification of Settlements [12].

To enable temporal and multivariate modelling, the following derived features were computed for each HPV genotype within each region–time stratum:

1. Prevalence (proportion positive). Computed monthly where temporal data permitted.

2. Infection intensity. Mean viral load (log_{10}-transformed copies/mL) among test-positive individuals within each stratum.

3. Moving averages. Three-month and six-month trailing moving averages of both prevalence and intensity, to smooth short-term fluctuations and highlight trend structure.

4. Lagged predictors. One-month and three-month lagged values of prevalence and intensity, capturing autoregressive temporal dependence.

5. Six-month trend direction. Binary indicator (positive or negative trend) computed from the sign of the linear regression slope fitted to the six most recent monthly observations.

A pairwise Pearson correlation matrix was computed across all genotype-level feature vectors (prevalence, intensity, moving averages, lagged values, and trend indicators), pooled across all region–time strata. Genotype pairs with fewer than 10 co-observed non-missing records were excluded from the matrix to ensure stable correlation estimates.

Hierarchical clustering of the resulting correlation matrix was performed using the *hclust* function in R 4.3.1 with average linkage (UPGMA) as the agglomeration criterion and $1 - r$ (where r is the Pearson correlation coefficient) as the distance metric. The number of genotype-level clusters was determined by visual inspection of the dendrogram supplemented by the Average Silhouette Width criterion in order to identify the co-expression modules of HPV genotypes and provide the genotype groupings used to characterize regional epidemiological risk profiles.

To identify epidemiological clusters at the regional level, k-means clustering was applied to the full feature matrix (prevalence, intensity, moving averages, lagged values, and trend indicators) for each region, using the *h2o* package in R 4.3.1. The optimal number of clusters (k) was determined by the Elbow Method applied to within-cluster sum of squares (WCSS) across $k = 2$ to $k = 8$, supplemented by the Gap Statistic. The selected value of $k = 4$ was confirmed by epidemiological face validity – each cluster yielded a coherent, interpretable regional profile consistent with known differences in urbanization, healthcare access, and genotype distribution.

K-means was initialized using the k-means++ seeding algorithm to reduce sensitivity to random initialization. The algorithm was run 100 times with different random seeds, and the solution with the lowest WCSS was retained. Cluster stability was assessed by bootstrapping (1.000 resamples of 60 % of records) and computing the Jaccard stability index for each cluster; clusters with Jaccard index > 0.75 were considered stable.

All analyses were conducted in R version 4.3.1. Key packages: dplyr (data manipulation), zoo (moving averages), stats::hclust (hierarchical clustering), h2o (k-means clustering), ggplot2 and pheatmap (visualisation).

Research results and their discussion

The International Agency for Research on Cancer (IARC) classifies HPV genotypes into three carcinogenic risk categories based on systematic reviews of epidemiological cohort data, mechanistic studies, and animal models. These categories – confirmed (Group 1), probable (Group 2A), and possible (Group 2B) carcinogens – determine clinical management pathways and screening triage protocols worldwide. The classification is not static: reclassification occurs as new evidence accumulates (e.g., HPV 68 was

reclassified from Group 2B to Group 2A). Awareness of this gradient is essential to correctly interpret regional genotype distribution data (Table 1).

Within the QVINTIP Recordati database analysed in this study, the following genotypes appeared across the four identified epidemiological clusters. Their precise IARC classification is provided below (Table 2) to enable direct mapping onto the colour-coded regional distribution shown in Fig. 1 and the co-expression structures depicted in Fig. 2 and 3.

Table 1

IARC oncogenic risk classification of clinically relevant HPV genotypes

IARC Classification	HPV Genotypes	Clinical Significance
High-risk (Group 1 – confirmed carcinogenic)	HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68	Strongly associated with cervical, oropharyngeal, anal, vulvar, and penile cancers. HPV 16 and 18 alone account for ~70 % of CC cases.
Probably carcinogenic (Group 2A)	HPV 68	Epidemiological and mechanistic evidence suggests carcinogenic potential; not yet confirmed as Group 1.
Possibly carcinogenic (Group 2B)	HPV 26, 53, 66, 67, 70, 73, 85	Limited evidence of carcinogenicity. May contribute to neoplasia in co-infection scenarios or immunocompromised hosts.
Low-risk (non-carcinogenic)	HPV 6, 11, 40, 42, 43, 44, 54	Cause genital warts (condylomata acuminata) and recurrent respiratory papillomatosis; rarely, if ever, progress to malignancy.

Table 2

Oncogenic risk category of HPV genotypes detected per cluster in the present study

Cluster	Region	Predominant HPV Types	Risk Category	CC Risk Level
1	Lutsk	No predominant types detected	N/A	Minimal
2	Odesa	HPV 11 (low-risk), HPV 51 (high-risk)	Mixed	Moderate
3	Zhyto-myr	HPV 16 (high-risk), HPV 44 (low-risk), HPV 66 (Group 2B)	Predominantly high-risk	High
4	Poltava	HPV 54 (low-risk), HPV 53 (Group 2B), HPV 35 (high-risk), HPV 58 (high-risk)	Mixed (high-risk dominant)	High

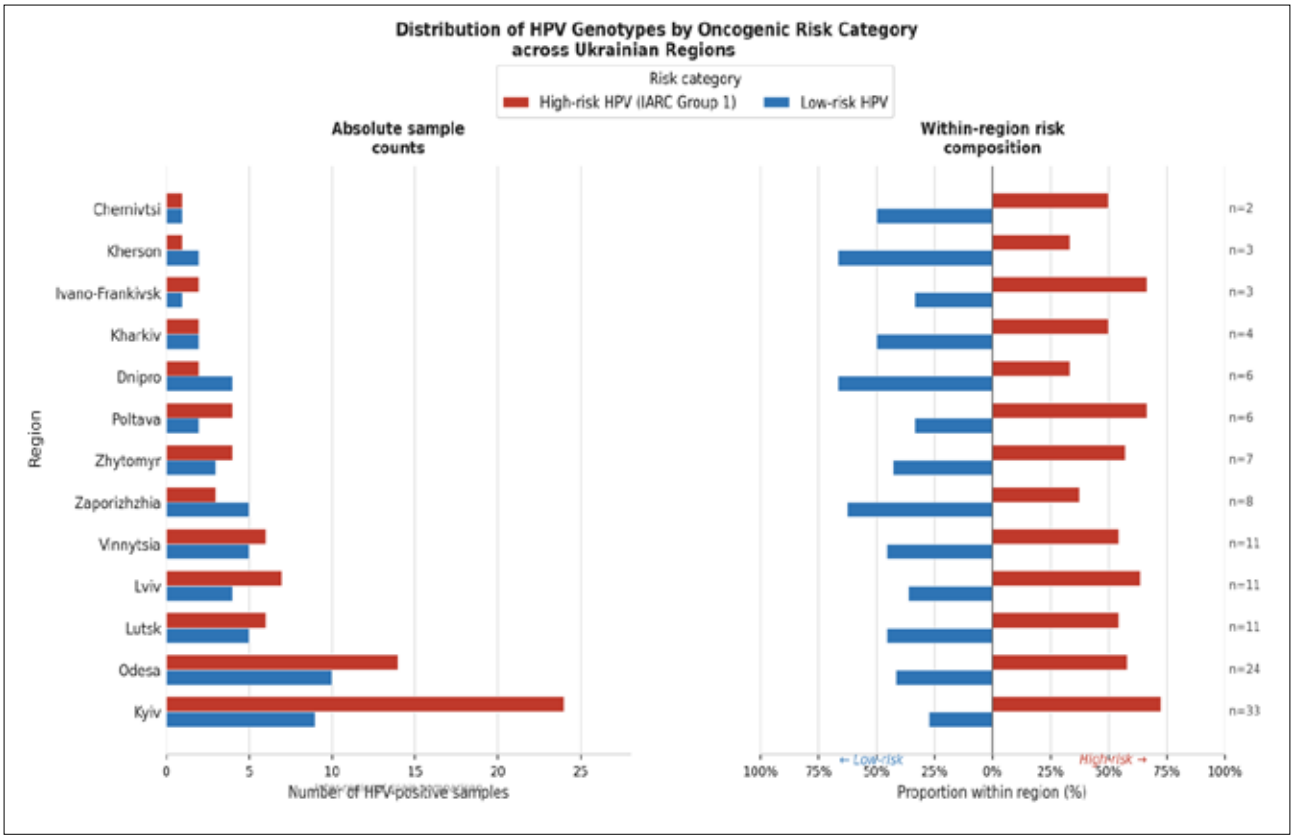


Fig. 1. Distribution of human papillomavirus genotypes with different oncogenic potential (by colour) in different regions (Performed by the authors based on the results of cluster analysis of QVINTIP Recordati data)

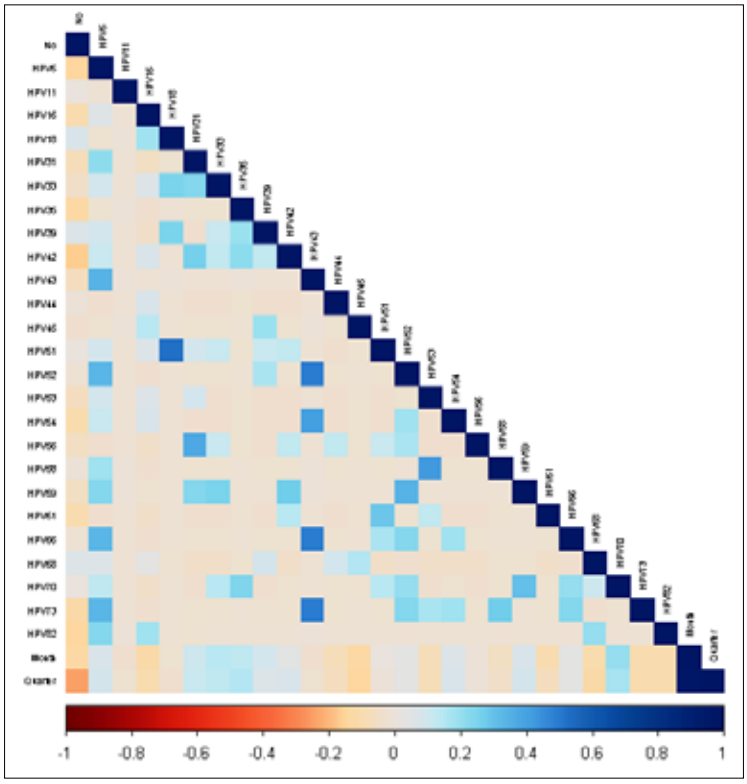


Fig. 2. Correlation matrix of accompanying oncogene expression
(Performed by the authors based on the results of cluster analysis of QVINTIP Recordati data)

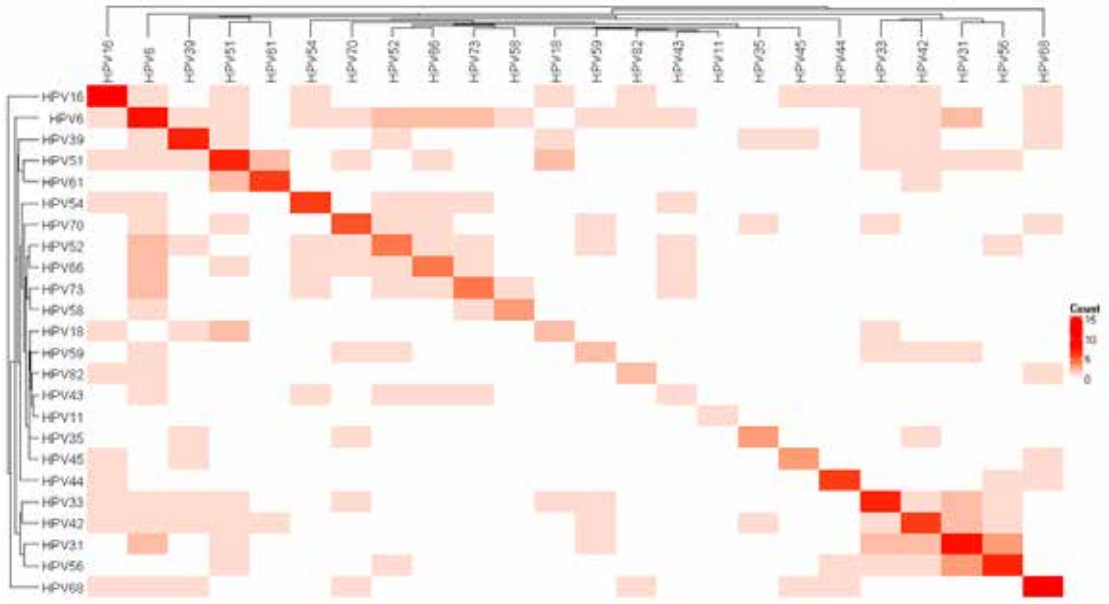


Fig. 3. Correlation heat map of accompanying oncogenes
(Performed by the authors based on the results of cluster analysis of QVINTIP Recordati data)

Fig. 1 demonstrates the geographical distribution of stratified HPV genotypes. It reveals a geographically unequal burden of high-risk genotypes. Larger urban centres (consistent with clusters 3 and 4, corresponding to Zhytomyr and Poltava regions) exhibit a preponderance of high-risk strains – notably HPV 16 in Zhytomyr and HPV 35/58 in Poltava – consistent with the finding that the detection rate of highly cancerogenic HPV strains is higher in urbanised regions. This urban concentration is biologically plausible: denser sexual networks, earlier sexual debut, and greater healthcare-seeking behavior (which increases detection) converge in urban settings.

The low-risk strain dominance in Cluster 1 (Lutsk) accounts for the minimal cervical cancer risk designation for that region. The apparent paradox of Cluster 4 (Poltava) – wherein the nominally low-risk HPV 54 is the

most frequently named type – is resolved by recognizing the co-occurrence of HPV 35 and HPV 58 (both Group 1 high-risk), which are present alongside HPV 54 within the same cluster profile. This co-infection context is precisely what frequency analysis (examining genotypes individually) usually misses.

Each cell of a pairwise Pearson correlation matrix of viral load and prevalence parameters across all monitored HPV genotypes (fig. 2) represents the strength and direction of co-variation between two genotype-level variables (prevalence, intensity, moving averages, or lagged predictors) across the regional time series.

High-risk genotypes cluster together in the upper-left quadrant of the matrix ($r > 0.6$), reflecting the tendency for Group 1 strains to co-occur – a well-established epidemiological phenomenon attributable to shared transmission routes, immunosuppressive interactions, and population-level sexual network overlap. Low-risk strains, vice versa, form a loosely correlated subcluster in the lower-right quadrant. The intermediate-risk strains (HPV 53, 66) occupy transitional positions in the correlation space, consistent with their intermediate carcinogenic classification. The matrix itself served a dual function: it provided the input distance structure for hierarchical clustering (Figure 3) and directly displayed which genotype combinations drive regional risk profiles – information that was primarily invisible from frequency analysis alone.

Figure 3 re-presents the correlation matrix after hierarchical cluster ordering – average linkage, and Euclidean distance on the correlation space. The heatmap reveals three primary co-expression modules: (i) a high-risk co-expression block centred on HPV 16, 31, 35, and 58; (ii) a moderate-risk block including HPV 51, 53, and 66; and (iii) a low-risk block including HPV 11, 44, and 54. The encoded correlation coefficient magnitude indicates strong positive correlation (genotypes that tend to rise and fall together across regions and observed time points). These modules are the molecular-epidemiological basis for the four epidemiological clusters described in the main text: regions in which the high-risk co-expression module is active (Zhytomyr, Poltava) receive elevated cervical cancer risk designations, while regions where only the low-risk module is expressed (Lutsk) receive minimal risk designations.

Frequency analysis – computing the proportion of positive tests for each HPV genotype within each geographic stratum – was the natural starting point for this investigation. It established the fundamental epidemiological parameters: an overall HPV prevalence of 38.5 %, and the observation that highly oncogenic strains appear more commonly in larger urban centres. These descriptive statistics are necessary for public health communication and for establishing baseline prevalence benchmarks comparable to published national and international estimates.

However, frequency analysis is inherently univariate and static: it examines each genotype in isolation and reports point-in-time proportions. It answers the question "how often does each genotype appear?" but cannot address "which genotypes co-occur?", "how do prevalence patterns relate to intensity trajectories over time?", or "which regions share similar multi-genotype epidemiological profiles?" – questions that are central to targeted clinical and public health responses.

Several specific limitations motivated the decision to proceed to multivariate analysis. First of all, frequency tables treat each HPV type as an independent variable. In reality, HPV genotypes co-infect the same host, compete for cellular niches, and interact immunologically. Ignoring these dependencies produces an incomplete – and potentially misleading – picture of infection burden and cervical cancer risk. A region with 20 % prevalence of HPV 54 (low-risk) and 15 % prevalence of HPV 35 (high-risk) has a very different risk profile from one with 35 % prevalence of HPV 54 alone; frequency tables cannot distinguish these scenarios. Secondly, the QVINTIP Recordati database includes longitudinal entries allowing computation of moving averages and lagged predictors. Frequency analysis collapses this time dimension into a single cross-sectional count, discarding evidence of emerging outbreaks, seasonal variation, and epidemiological trends that are crucial for forecasting and for evaluating screening interval adequacy. Thirdly, effective public health policy requires identifying regions with similar epidemiological profiles to design scalable, group-specific interventions. With 195 records spanning multiple genotypes and regions, pairwise comparison of all regional frequency tables is computationally and interpretively intractable. Clustering reduces this complexity to a manageable set of archetypes. And finally, the feature space constructed from genotype-level prevalence, intensity, moving averages, lagged values, and trend indicators is high-dimensional. Standard frequency tables cannot account for multicollinearity among these features, inflating apparent differences between regions that arise from correlated rather than independent sources of variation.

When applied, the pairwise correlation matrix of genotype-level features serves two distinct analytical purposes – biological insight and input for distance-based clustering. First, it revealed which genotypes co-varied across regions and time, identifying co-expression modules that reflected shared epidemiological determinants (e.g., sexual network overlap, shared immune evasion mechanisms). These modules, rather than individual genotype frequencies, constituted the biologically meaningful units of regional HPV epidemiology. Secondly, the correlation matrix defined a similarity structure among genotype-level variables. Hierarchical clustering applied to this matrix grouped genotypes into co-expression blocks without imposing a pre-specified number of clusters, outperforming non modular data assumption approaches.

For genotype-level co-expression structure, the optimal k is not epidemiologically obvious, and imposing an arbitrary k risks both over-splitting coherent modules and merging biologically distinct ones, let alone non-parametric characteristic of the dataset. Hierarchical clustering circumvents this limitation: it produces a full dendrogram from which any number of clusters can be cut, allowing the analyst to inspect cluster stability across the full hierarchy before committing to a particular cut-point.

Overall, this approach aligns with established practices in infectious disease epidemiology when analyzing multi-pathogen or multi-genotype surveillance data, and

analogous approaches have been applied in influenza subtype co-circulation analysis and HIV phylodynamic studies.

The overall HPV seropositivity rate of 38.5 % observed in this study substantially exceeds the WHO-estimated global average prevalence of approximately 11–12 % in women with normal cytology, and is nearly double the European average of approximately 15–20 % reported in systematic reviews. This discrepancy requires careful contextualisation to avoid misinterpretation. So, the QVINTIP Recordati database is a clinical registry rather than a population-based random sample. Women entered into the database are, by definition, individuals who presented for gynaecological examination or screening – a population systematically enriched for symptomatic, high-risk, or screening-eligible individuals. Selection bias of this nature is well-recognised in HPV surveillance and inflates observed prevalence relative to true population-level estimates. Accordingly, the 38.5 % figure should be interpreted as the prevalence within a clinically presenting population, not as an estimate of nationwide community-level seroprevalence.

Second, the comparison benchmark of "twice the European level" is nonetheless diagnostically significant, because the most directly comparable European clinical-sample registries – which share the same selection mechanism – report prevalences in the 18–24 % range. The excess above even these clinically-biased comparators reflects genuine elevated burden attributable to inadequate vaccination coverage and screening infrastructure in Ukraine. The Ukrainian national HPV vaccination rate remains below 5 % – among the lowest in Europe – and opportunistic rather than systematic screening is the norm, meaning that many infections progress undetected until advanced-stage disease.

Third, the large-scale Ukrainian study spanning 2014–2016 [12], which analysed more than 40.000 women and reported an overall HPV detection rate of 34 %, provides an important internal benchmark. Our findings of 38.5 % is directionally consistent, with the modest upward difference attributable to the smaller, more clinically-selected sample and potential temporal changes in genotype circulation patterns. This is consistent with international evidence but requires mechanistic unpacking to be actionable for policy.

Urbanised populations have substantially higher rates of gynaecological attendance and screening utilisation. The observed higher prevalence of high-risk genotypes in cities may, in part, reflect superior case ascertainment rather than a true biological excess of infection. Independently of detection bias, biological mechanisms may explain a genuine urban excess of high-risk strains. Larger urban sexual networks exhibit shorter mean network distances between individuals, accelerating genotype spread. High-risk genotypes (particularly HPV 16 and 18) have demonstrated evolutionary fitness advantages in densely connected sexual networks, as they are more efficiently transmitted per sexual contact than low-risk types and exhibit superior immune evasion characteristics. The relative preponderance of HPV 16 in Zhytomyr (a major regional city) is consistent with this mechanism.

Urban populations exhibit earlier age of sexual debut and higher numbers of lifetime sexual partners on average, both of which are robust HPV risk factors. However, urban women also have higher educational attainment and income, which are associated with greater uptake of preventive services – creating counterbalancing forces that likely explain why the urban–rural gradient in Ukraine is present but not extreme.

The consistently low HPV activity and absence of predominant high-risk genotypes in Cluster 1 make it the highest priority for *preservation-focused intervention*: maintaining low prevalence through vaccination and health education rather than intensive screening. Universal HPV vaccination of pre-adolescent girls (9–14 years) before sexual debut would be most cost-effective in this setting, capitalising on the currently low baseline to achieve herd immunity before high-risk strains establish endemic circulation.

The co-presence of HPV 11 and HPV 51 – a low-risk and a high-risk strain, respectively – in the Odesa cluster is epidemiologically significant because mixed-risk profiles complicate clinical triage. HPV 11 infection, while non-carcinogenic, causes recurrent respiratory papillomatosis and genital condylomata, imposing significant morbidity and healthcare costs. The concurrent HPV 51 burden requires enhanced cytological surveillance. Localised outbreak dynamics noted in Cluster 2 (fluctuating prevalence over the six-month observation window) suggest episodic introduction events – potentially consistent with population mixing via maritime trade – and argue for real-time surveillance rather than annual cross-sectional screening.

The dominance of HPV 16 in Zhytomyr – a genotype responsible for approximately 50–60 % of all cervical cancers globally – combined with the co-occurrence of HPV 66 (Group 2B) and HPV 44 (low-risk, providing a marker of sexual activity) constitutes the highest cervical cancer risk profile in the dataset. HPV 16 has the longest latency from infection to invasive carcinoma (mean approximately 15 years) and the strongest propensity for chromosomal integration and E6/E7 oncoprotein-mediated p53/Rb pathway disruption. The policy implication is unambiguous: Zhytomyr requires immediate prioritisation for both bivalent/quadrivalent HPV vaccination and primary HPV DNA testing-based screening, with cytological triage for HPV 16/18-positive cases.

Cluster 4 presents an analytically complex pattern: high overall HPV activity dominated by HPV 54 (low-risk) alongside HPV 53 and HPV 66 (both Group 2B, possibly carcinogenic) and confirmed high-risk HPV 35 and HPV 58. This constellation is not well-characterised in the existing literature, and the mechanisms driving the regional predominance of HPV 54 – a genotype not typically associated with epidemic spread – are unknown. Possible explanations include a founder effect from a specific sexual network subpopulation, assortative sexual mixing patterns, or laboratory detection artefact. An in-depth molecular epidemiological investigation (phylogenetic analysis of local HPV 54 isolates, sexual network mapping) is warranted before designing region-specific interventions.

All four cluster profiles would be directly addressed by a systematic national HPV vaccination programme [13; 14].

Ukraine's vaccination coverage fell precipitously following conflict-related disruption to healthcare infrastructure and cold-chain logistics, reaching levels below 5 % of the target cohort. In contrast, countries that have maintained multi-year systematic vaccination – notably the United Kingdom (achieving >80 % first-dose coverage in girls aged 12–13), Sweden, and Australia – have documented 50–80 % reductions in vaccine-targeted genotype prevalence within a decade of programme initiation.

Crucially, the cross-protective benefit of bivalent (HPV 16/18) and nonavalent (HPV 6, 11, 16, 18, 31, 33, 45, 52, 58) vaccines extends beyond the directly targeted genotypes. Post-vaccination cohort studies consistently document partial cross-protection against HPV 31, 33, and 45 – genotypes detectable in the Zhytomyr and Poltava clusters [14]. This cross-protection effect would be expected to reduce the prevalence of all high-risk types identified in the present study, not merely the vaccine-targeted HPV 16 and 18.

Interestingly, the economic case for vaccination in Ukraine is compelling. The direct medical costs of treating CIN II/III and invasive cervical cancer – including hospitalization, surgery, chemotherapy, and radiotherapy – vastly exceed the per-dose vaccine cost at current international procurement prices, particularly when productivity losses among women of reproductive age are included in the economic model. Extending vaccination to catch-up programs for women up to age 26, and to boys (reducing HPV transmission and HPV-associated oropharyngeal cancer burden), would further amplify cost-effectiveness.

However, several limitations constrain the generalisability of this work and motivate future investigations. The narrow dataset, while sufficient for exploratory cluster analysis, is objectively underpowered for stable estimation of genotype-specific prevalence at the regional level. Cluster center-point estimates should be treated as provisional pending validation in a larger, population-representative cohort. Also, the QVINTIP Recordati registry is a clinical registry with inherent selection bias towards symptomatic or high-risk presenting populations. Population-based prevalence estimation would require random sampling from the general female population, stratified by age, region, and urbanisation grade. Although moving averages and lagged predictors were computed from longitudinal entries, the temporal coverage of the dataset did not provide any enlightening results. Therefore, the observations clustered within a narrow time window, the temporal modelling assumptions alone were not justified for report.

Additionally, age, parity, contraceptive history, smoking status, and HIV serostatus are established HPV risk factors of a potential factor model whose absence from the clustering feature set alone limits the explainability of the identified clusters. Further investigation should integrate these covariates as both clustering features and outcome modifiers. Finally, data collection occurring during or proximate to active armed conflict in Ukraine may have introduced systematic biases in healthcare-seeking behavior, regional accessibility, and laboratory quality control – all of which could alter reported prevalences.

Conclusions

1. The overall prevalence of HPV among clinically presented women in Ukraine was 38.5 %, substantially exceeding average European levels, highlighting a high burden of infection in the country. High-risk genotypes, particularly HPV-16 and HPV-18, which are strongly associated with cervical cancer development, were most frequently detected.
2. The results of the cluster analysis confirmed that urbanized regions demonstrated an increased prevalence of high-risk HPV genotypes compared to less populated areas, which can be explained by a combination of biological, behavioral (high population density), and diagnostic (optimized access to screening programs) factors.
3. Multivariate analysis of HPV genotype co-expression identified four epidemiological clusters, representing distinct regional risk profiles for cervical cancer. Clusters dominated by high-risk strains (Zhytomyr, Poltava) exhibited elevated risk for CIN III and invasive cervical cancer, whereas clusters with low HPV activity (Lutsk) were associated with minimal risk.
4. Coinfections and co-expression modules indicate that simple frequency counts of individual HPV genotypes do not fully capture risk patterns [15]. A multivariate approach enables a more accurate assessment of regional epidemiological profiles and screening priorities.
5. Inadequate HPV vaccination coverage and low participation in screening programs underscore the need to strengthen preventive strategies, including universal vaccination and educational interventions, particularly in regions with a high burden of high-risk genotypes.
6. Further studies on representative populations, integrating sociodemographic factors, behavioral risks, and clinical characteristics, are essential to optimize regionally tailored strategies for HPV infection prevention and cervical cancer control in Ukraine.

REFERENCES

1. Fedorenko ZP, Sumkina OV, Gorokh YeL, Goulak LO, Ryzhov AYu. Cancer in Ukraine 2022–2023: Incidence, mortality, prevalence and other relevant statistics. *Bull Nat Cancer Regist Ukr*. 2024;25. Available from: https://www.ncru.inf.ua/publications/BULL_25/index_e.htm.
2. Bondar OV, Rybin AI, Muzyka VV. Cervical cancer screening programme with HPV testing. *Bull Marit Med*. 2022;(1(94)):58–64. doi: https://doi.org/10.5281/zenodo.6414037.
3. Bondar OV, Rybin AI, Muzyka VV. Experience of primary HPV testing in cervical cancer screening. *Odesa Med J*. 2022;(1–2):93–96. doi: https://doi.org/10.54229/2226-2008-2022-1-2-16.
4. Gray P, Wang J, Nordqvist Kleppe S, Elfström KM, Dillner J. Population-based age-period-cohort analysis of declining human papillomavirus prevalence. *J Infect Dis*. 2025;231(4):e638–e649. doi: https://doi.org/10.1093/infdis/jiaf032.
5. Zahnd WE, Rodriguez C, Jenkins WD. Rural-urban differences in human papillomavirus-associated cancer trends and rates. *J Rural Health*. 2019;35(2):208–215. doi: https://doi.org/10.1111/jrh.12305.

6. Mariño JM, Rocha DAP, Reis RS, et al. HPV-positive women living in isolated areas in Amazonas, Brazil: clinical-epidemiological profile and cytological findings. *Cad Saude Colet*. 2024;32(1):e32010365. doi: https://doi.org/10.1590/1414-462X202432010365.
7. Zhao X, Shen S, Su C, Chang J, Yan Y, Zhao J. Epidemiological study of HPV infection in 24,588 rural women in Luonan, China. *Gynecol Oncol Rep*. 2024;57:101669. doi: https://doi.org/10.1016/j.gore.2024.101669.
8. Ehman KM, Jenkins GD, Grimm JA, et al. Primary human papillomavirus test uptake and cervical cancer screening trends in the Midwest, United States. *J Prim Care Community Health*. 2024;15:21501319241251934. doi: https://doi.org/10.1177/21501319241251934.
9. Zhetpisbayeva I, Kassymbekova F, Sarmuldayeva S, Semenova Y, Glushkova N. Cervical cancer prevention in rural areas. *Ann Glob Health*. 2023;89(1):75. doi: https://doi.org/10.5334/aogh.4133.
10. Rybin A, Liskovskyi S. Epidemiology of papillomavirus infection in Ukraine. *Pedag Psychol Sport*. 2025;26:66875. doi: https://doi.org/10.12775/PPS.2025.26.66875.
11. International Agency for Research on Cancer. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Vol. 100B, Biological Agents. Lyon, France: IARC; 2012.
12. Law of Ukraine “On the Procedure for Resolving Certain Issues of the Administrative Territorial Structure of Ukraine” No. 3285-IX, 28 Jul 2023. Available from: https://zakon.rada.gov.ua/laws/show/3285-20#Text.
13. World Health Organization. Ukraine adds HPV vaccine to the national immunization schedule. 2026. Available from: https://www.who.int/ukraine/news/item/10-03-2026-ukraine-adds-hpv-vaccine-to-the-national-immunization-schedule.
14. de Bondt D, Naslazi E, Jansen E, et al. Validating the predicted impact of HPV vaccination on HPV prevalence, cervical lesions, and cervical cancer: A systematic review of population level data and modelling studies. *Gynecol Oncol.* 2025;195:134–143. doi: https://doi.org/10.1016/j.ygyno.2025.03.008.
15. Zhou D, Xue J, Sun Y, et al. Patterns of single and multiple HPV infections in female: A systematic review and meta-analysis. *Heliyon*. 2024;10(17):e35736. doi: https://doi.org/10.1016/j.heliyon.2024.e35736.

Надійшла до редакції 25.09.2025 р.
Прийнята до друку 29.06.2026 р.
Електронна адреса для листування viktoria.muzyka.m1@gmail.com