Multiple high-risk HPV genotypes are grouped by type and are associated with viral load and risk factors

L. DEL RÍO-OSPINA[1,2], S. C. SOTO-DE LEÓN[1,3], M. CAMARGO[1,4],
R. SÁNCHEZ[1,2], D. A. MORENO-PÉREZ[1,4], A. PÉREZ-PRADOS[5],
M. E. PATARROYO[1,2] AND M. A. PATARROYO[1,6]*

[1] *Molecular Biology and Immunology Department, Fundación Instituto de Inmunología de Colombia (FIDIC), Bogotá D.C., Colombia*
[2] *School of Medicine, Universidad Nacional de Colombia, Bogotá D.C., Colombia*
[3] *Universidad de Ciencias Aplicadas y Ambientales (UDCA), Bogotá D.C., Colombia*
[4] *PhD Programme in Biomedical and Biological Sciences, Universidad del Rosario, Bogotá D.C., Colombia*
[5] *Mathematics Department, Universidad Pública de Navarra, Pamplona, Spain*
[6] *School of Medicine and Health Sciences, Universidad del Rosario, Bogotá D.C., Colombia*

SUMMARY

Investigating whether high-risk human papillomavirus (HR-HPV) types tend to become grouped in a particular way and whether factors are associated with such grouping is important for measuring the real impact of vaccination. In total, 219 women proving positive for HPV as detected by real-time PCR were included in the study. Each sample was analysed for detecting and quantifying six viral types and the hydroxymethylbilane synthase gene. Multiple correspondence analysis led to determining grouping patterns for six HR-HPV types and simultaneous association with multiple variables and whether viral load was related to the coexistence of other viral types. Two grouping profiles were identified: the first included HPV-16 and HPV-45 and the second profile was represented by HPV-31, HPV-33 and HPV-58. Variables such as origin, contraceptive method, births and pregnancies, educational level, healthcare affiliation regime, atypical squamous cells of undetermined significance and viral load were associated with these grouping profiles. Different socio-demographic characteristics were found when coinfection occurred by phylogenetically related HPV types and when coinfection was due to non-related types. Biological characteristics, the number of viral copies, temporality regarding acquiring infection and competition between viral types could influence the configuration of grouping patterns. Characteristics related to women and HPV, influence such interactions between coexisting HPV types reflecting the importance of their evaluation.

Key words: Grouping, high-risk human papillomavirus, multiple infection, risk factors, viral DNA load.

INTRODUCTION

Persistent infection by at least one type of high-risk human papillomavirus (HR-HPV) is necessary but not sufficient for such infection to progress to cervical cancer (CC) [1]. Bearing in mind that the main HPV

* Author for correspondence: M. A. Patarroyo, Molecular Biology and Immunology Department, Fundación Instituto de Inmunología de Colombia (FIDIC), Cra 50 # 26-20, Bogotá, Colombia.
(Email: mapatarr.fidic@gmail.com)

transmission route is sexual, females infected by at least one viral type have a greater probability of acquiring new viral types [2–4].

Variable coinfection prevalence has been reported [5–7] and it has been found that the number of cases of multiple HPV infection has been significantly greater than expected [5, 8]. It has been described that infection by multiple types of HPV plays an important role in the development and/or progression of precursor CC lesions [5, 9]. The inverse relationship between coinfection prevalence and cervical lesion severity [7] has led to exploring which HPV genotypes tend to be associated with coinfection with which other types could be useful for understanding the impact of multiple infection and identifying females at risk of developing CC.

Knowing whether determined groupings of HR-HPV types exist in multiple infection at phylogenetic or genotype level has been of interest for some investigations where different populations and detection techniques have been used, unfortunately leading to non-conclusive findings so far [5, 10]. Moreover, limitations related to primer sets leading to underestimating the number of types [9], as well as a restricted sample size and the need for multiple statistical analyses [8], difficult interpreting grouping patterns.

The present study has been aimed at investigating whether multiple HPV types (the six most prevalent high-risk types in Colombia [11]), particularly certain combinations of types, tend to become grouped and which factors (related to women and HPV) are associated to such grouping patterns. This could provide additional information for understanding the possible impact of multiple viral types regarding carcinogenesis and thus establish a baseline for future studies aimed at monitoring the impact of vaccination on a target population.

METHODS

Study population

All the females who participated in this study voluntarily attended the existing CC promotion and prevention (P&P) programmes being offered by healthcare institutions in three Colombian cities (Bogotá, Girardot and Chaparral). The methodology used in this study has previously been described [12].

Females in whose cervical samples infection was detected involving at least one type of HR-HPV from the six types identified by real-time polymerase chain reaction (PCR) and with complete medical history information were included for analysis in this study ($n = 242$). Patients were excluded in whose samples the constitutive gene hydroxymethylbilane synthase (HMBS) could not be amplified ($n = 23$). The sample thus consisted of 219 females.

Each participating institution's ethics committee approved the study and supervised it (i.e. the Ethics Committees at Fundación Instituto de Inmunología de Colombia, Nuevo Hospital San Rafael E.S.E in Girardot, the Hospital San Juan Bautista de Chaparral E.S.E. and the Hospital de Engativá).

Detecting and quantifying DNA from HPV

The methodology used in this study has previously been described by our research group [12]. Briefly, DNA from the six high-risk viral types previously reported as having high frequency in Colombia was detected and quantified by real-time PCR. The 1:10 serial dilution (10^{11} to 10^6) was obtained having an initial concentration of plasmid for each HPV type and HMBS. These six dilution points were used for standardising real-time PCR by constructing standard curves for each viral type and the HMBS gene. Each sample was analysed for detecting and quantifying HPV-16, HPV-18, HPV-31, HPV-33, HPV-45 and HPV-58. The HMBS gene was amplified to verify DNA integrity and determine the number of viral copies per cell. Four parallel duplex real-time PCR were run for each patient (the first for HPV-16; the second included HPV-18 and -31; the third for HPV-33 and -45; and the last one for HPV-58 and HMBS), using the CFX96 Touch (BIO-RAD) detection system. Six plasmid dilution points were included in each analysis for respective viral types or HMBS as well as an NTC (no template control). Absolute viral load and normalised viral load (HPV copies/cell = number of HPV copies/(number of HMBS copies/2)) [13] were thus quantified for each viral type.

Statistical analysis

Frequencies and percentages were used for univariate analysis of categorical variables; medians and interquartile ranges (IQR) were used for quantitative variables, bearing the non-normal distribution of data in mind. The presence or absence of infection for each HPV type was established by viral load, using >0 viral copies as cut-off point for defining viral load positivity.

Multiple infection status was defined as real-time PCR detection of two or more viral types (coinfection). It was determined whether there were differences regarding socio-demographic variables and risk factors between females suffering single infection and those suffering coinfection. Fisher's exact test was used for identifying such differences. The Mann–Whitney test was used for quantitative variables, evaluating differences in distribution between both groups of females. STATA 10 software was used for analysing data; all hypothesis tests were two-tailed, using 0·05 significance.

Logistic regression analysis was used for modelling multiple infection risk. The variables included in the model were age, origin, ethnicity, a background of smoking cigarettes, age on first sexual relationship, lifetime number of sexual partners, contraceptive method used, a background of births and previous abortions, menopause and STD (sexually transmitted disease). A Poisson regression model was used for evaluating the variables associated with the amount of infecting types (1–6 types) regarding count data; this included the same set of variables as in the logistic model. The assumption of equivalence between variance and expected value regarding these models had been previously verified through over-dispersion tests.

Multiple correspondence analysis (MCA) was used for exploring HPV genotype grouping. MCA is a form of factor analysis aimed at grouping a set of variables into a smaller set of related factors or profiles. Such analysis evaluated the degree to which each of the six HPV types participated in configuring determined profiles and the simultaneous association between multiple variables within the profiles so detected (coinfection patterns, cytological findings and their relationship with some variables considered risk factors for HPV infection). This method enabled analysing the profiles configured by multiple variables in terms of different modalities' similarity or proximity [14]. The significant representation of each variable was taken into account when selecting variables to avoid results being due to chance (as in other types of multivariate analysis) [14, 15]. This led to identifying groups having clinical significance from different groupings of categories of variables and continuous variables' patterns, i.e. using this method led to identifying how HPV types were grouped and how socio-demographic characteristics, risk factors and cytology result were associated with determined groups of viral types.

Two groups of variables were chosen in this study for such analysis: active variables used in constructing factorial axes and supplementary or illustrative variables enriching interpretation of factorial axes once they have been constructed [15]. The presence/absence of infection caused by each viral type was considered an active variable for this analysis, whilst some HR-HPV infection risk factors, cytological findings and viral load for each HR type were treated as illustrative variables.

The factorial axes on which the grouping of variables had to be evaluated were initially established by analysing the variance and eigenvalues in the histogram constructed from active variables (i.e. infection by each of the six viral types detected). Two factorial axes were selected, bearing six initial variables in mind (HPV-16, −18, −31, −33, −45 and −58), the percentage of variance explained by each variable and the variance accumulated by the first two factors (close to 50%). Furthermore, contribution values were analysed for interpreting these axes, thereby establishing which categories were meaningful within each axis. The average contribution of all categories of active variables was thus determined (each of six variables had two categories: yes and no; $(100/12 = 8·3\%)$), categories making more than 8·3% contribution were selected.

Cosine squared values were also evaluated to complement the contribution values, thereby leading to an estimation of the quality of the representation of each active variable on each axis (which was better as it was closer to 1) and test values were evaluated for determining whether the representation of each viral type within a determined grouping of viral types (axis) was significantly different from 0, giving an assessment of each variable's significance.

Selecting illustrative variables involved independently analysing the representation of each category for each variable on each factorial axis (or groups of HPV types). Test values were thus taken into account, indicating whether the location coordinates for each category on the two-dimensional plane (determined by both factors) was significantly different from 0, thereby ascertaining whether the representation of each category in a factor was significant, taking $\leqslant -2$ or $\geqslant 2$ values as cut-off points (determined by MCA). This led to identifying categories having $\leqslant -2$ or $\geqslant 2$ test values, thereby determining the variables having the greatest representation in the grouping of HPV types. Variables associated with a particular HPV group were identified according to

the test value sign (negative or positive). Categories having ⩽−2 test values were associated with the grouping found at the negative poles for each factor; categories having ⩾2 test values were associated with the grouping found at the positive poles. The final model only included variables having test values establishing significant representation (⩽−2 or ⩾2) for a particular group of HPV types.

Viral loads were included as continuous illustrative variables. Their representation on each axis was analysed bearing correlation values in mind.

The different profiles' formation and structure were analysed using a two-dimensional graphical representation giving the socio-demographic variables, risk factors and cytological findings' projection and each viral type's contribution, as well as the correlation of the six HR-HPV viral loads for each profile configured here. Bearing in mind the distance between the variables represented on the graph, possible dependence and similarity relationships were identified regarding the categories so represented. SPAD-5 software was used for all multiple correspondence statistical analysis.

Absolute and normalised viral loads were analysed; however, only absolute load was used for multivariate analysis (MCA) since this allowed a closer estimation of the total amount of viral DNA. Absolute load values were transformed in log10 to facilitate interpreting the results.

RESULTS

Socio-demographic characteristics, risk factors and cervical findings

The study population's median age was 43 years (IQR 37–51 years). No statistically significant differences were found when comparing the medians for female age between those with single and multiple infections ($P = 0.308$).

The prevalence of socio-demographic characteristics, risk factors and cytology and colposcopy findings was determined in the study population (Table 1); however, no statistically significant differences were found when comparing both groups of patients.

Even though only five females were diagnosed as having atypical squamous cells of undetermined significance (ASCUS), all of them were positive for infection by more than one viral type. There was greater prevalence of coinfection in females having negative cytology and colposcopy than those with squamous intraepithelial lesions (SIL).

HPV infection and viral load

Multiple infection prevalence was 75·80% ($n = 166$). HPV-18 was the most frequently occurring viral type for both single ($n = 17$, 32·08%) and multiple infections ($n = 114$, 68·67%), followed by HPV-16 in multiple infections and HPV-31 regarding single infection (Table 2). Statistically significant differences were seen when comparing each viral type's frequency regarding single and multiple infections ($P = 0.000$), except for HPV-33 ($P = 0.079$).

Table 2 gives the medians for viral loads for each HR-HPV type. Statistically significant differences were only found for HPV-31 absolute viral load when determining whether there were differences regarding viral load distribution according to the presence of single or multiple infections, this being greater regarding coinfection ($P < 0.05$). It was seen that viral load was greater for HPV-31 and HPV-33 regarding multiple infections whilst this became reduced regarding HPV-58.

Multivariate analysis (logistic regression, Poisson and MCA models)

A multivariate logistic regression taking as outcome the infection by more than one HPV type was carried out; no significant results were found for any of the variables considered (data not shown). A Poisson regression using the amount of infecting HPV types as dependent variable was also performed. No significant results were found for any of the variables considered in the univariate analysis, so an MCA was thus used for exploring HPV genotype grouping and the associated variables.

Forming factorial axes

MCA led to identifying two factorial axes, thereby enabling the grouping of variables into two main profiles. When reviewing each active variable's contribution, cosine squared and test values (Table 3), it was found that presence of HPV-16 and HPV-45 were grouped on the positive pole of axis 1 and the absence of infection by these types was representative on the negative pole of axis 1. Regarding axis 2, HPV-33 and HPV-58 made the highest contribution and had the highest test and cosine squared values for this factor. The foregoing was also observed for HPV-31,

Table 1. *Socio-demographic characteristics, risk factors and cervical findings for the population being studied*

Characteristic	*n*	%	Single infection (*n* = 53)		Multiple infection (*n* = 166)		*P*
			n	%	*n*	%	
Age, in years							
⩽ 30	30	13·70	8	15·09	22	13·25	0·627
31–40	58	26·48	15	28·30	43	25·90	
41–50	75	34·25	20	37·74	55	33·13	
>50	56	25·57	10	18·87	46	27·71	
Origin							
Bogotá	73	33·33	19	35·85	54	32·53	0·894
Chaparral	5	2·28	1	1·89	4	2·41	
Girardot	141	64·38	33	62·26	108	65·06	
Ethnicity							
Mestizo	205	93·61	51	96·23	154	92·77	0·848
White	12	5·48	2	3·77	10	6·02	
Black	1	0·46	0	0·00	1	0·60	
Indigenous	1	0·46	0	0·00	1	0·60	
Average monthly income*							
⩽ minimum	192	87·67	47	88·68	145	87·35	0·506
>minimum	27	12·33	6	11·32	21	12·65	
Educational level							
Illiterate	5	2·28	1	1·89	4	2·41	0·530
Primary	107	48·86	26	49·06	81	48·80	
Secondary	88	40·18	20	37·74	68	40·96	
Technical qualification	11	5·02	2	3·77	9	5·42	
Graduate	8	3·65	4	7·55	4	2·41	
Marital status							
Married	29	13·24	6	11·32	23	13·86	0·856
Living with partner	158	72·15	41	77·36	117	70·48	
Widow	2	0·91	0		2	1·20	
Divorced	10	4·57	1	1·89	9	5·42	
Single	20	9·13	5	9·43	15	9·04	
Healthcare scheme affiliation							
Private	14	6·39	2	3·77	12	7·23	0·450
Contributory	5	2·28	0	0	5	3·01	
Subsidised	139	63·47	38	71.70	101	60·84	
Linked	61	27·85	13	24·53	48	28·92	
Smoker							
No	176	80·37	47	88·68	129	77·71	0·056
Yes	43	19·63	6	11·32	37	22·29	
Age at first intercourse, in years							
⩽ 18	129	58·90	27	50·94	102	61·45	0·117
>18	90	41·10	26	49·06	64	38·55	
Lifetime number of sexual partners							
1	104	47·49	25	47·17	79	47·59	0·836
2–3	98	44·75	23	43·40	75	45·18	
>3	17	7·76	5	9·43	12	7·23	
Contraceptive method							
None	89	40·64	26	49·06	63	37·95	0·654
Surgery	70	31·96	14	26·42	56	33·73	
Hormonal	23	10·50	4	7·55	19	11·45	
IUD	27	12·33	7	13·21	20	12·05	
Barrier	10	4·57	2	3·77	8	4·82	

Table 1 (*cont.*)

Characteristic	n	%	Single infection (n = 53)		Multiple infection (n = 166)		P
			n	%	n	%	
Pregnancies							
None	5	2·28	2	3·77	3	1·81	0·444
1–2	82	37·44	22	41·51	60	36·14	
3–4	100	45·66	20	37·74	80	48·19	
>4	32	14·61	9	16·98	23	13·86	
Births							
None	6	2·74	2	3·77	4	2·41	0·684
1–2	109	49·77	29	54·72	80	48·19	
3–4	92	42·01	19	35·85	73	43·98	
>4	12	5·48	3	5·66	9	5·42	
Abortions							
No	96	43·84	21	39·62	75	45·18	0·292
Yes	123	56·16	32	60·38	91	54·82	
Menopause							
No	171	78·08	46	86·79	125	75·30	0·054
Yes	48	21·92	7	13·21	41	24·70	
STD							
No	178	81·28	43	81·13	135	81·33	0·559
Yes	41	18·72	10	18·87	31	18·67	
Cytological findings							
Negative	205	93·61	48	90·57	157	94·58	0·054
ASCUS	5	2·28	0	0·00	5	3·01	
LSIL	9	4·11	5	9·43	4	2·41	
Histological findings							
Negative	158	72·15	36	67·92	122	73·49	0·463
LSIL	58	26·48	17	32·08	41	24·70	
HSIL	3	1·37	0	0·00	3	1·81	

P, *P* value; IUD, intrauterine device; STD, sexually transmitted disease; ASCUS, atypical squamous cells of undetermined significance; LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion.
* The minimum average monthly income (2014 rate) would be roughly US$ 300.

which was also representative for this axis; even when equal cosine squared values were obtained on both axes and higher test values on axis 1, the contribution values were significant for axis 2 but not for axis 1 (10·90 and 11·62 > 8·3% (average contribution)) (Table 3). It was found that the presence of HPV-31, HPV-33 and HPV-58 was most representative for the negative pole, whilst the absence of infection by HPV-31 and HPV-58 was most representative for the positive pole of axis 2.

Regarding HPV-18, it was found that contribution and cosine squared values, as well as correlation coefficients were not as high as those for other viral types, thereby indicating this variable's low representation on both axes, i.e. it was not significant for the profiles of groupings configured by the other viral types.

Grouping nominal and continuous illustrative variables on factorial axes (profiles)

Correlation coefficients were obtained for viral loads on each axis (Table 3). HPV-16 and HPV-45 viral loads had the highest coefficients for axis 1, thereby agreeing with that found regarding HPV-16, HPV-31 and HPV-45 representativeness on axis 1. Likewise, HPV-33 and HPV-58 viral loads had higher coefficients for axis 2.

The first profile grouped HPV-16 and HPV-45 (from different species: A9 and A7, respectively), having high viral loads (represented by the length and projection of the vectors). The variables associated with HPV-16 and HPV-45 (bearing significant test values in mind) were Girardot, background of 3–4

Table 2. *HPV genotype and absolute viral load distribution in single and multiple infections*

Viral type	n = 219				Single infection (n = 53)				Multiple infection (n = 166)				P value*
	% (n)	Median	p25	p75	% (n)	Median	p25	p75	% (n)	Median	p25	p75	
HPV-16	49·77 (109)	6·50	5·69	7·09	16·98 (9)	6·50	4·80	7·63	60·24 (100)	6·51	5·76	7·09	0·991
HPV-18	59·82 (131)	6·94	6·29	8·08	32·08(17)	7·04	6·27	1·51	68·67 (114)	6·90	6·30	8·10	0·639
HPV-31	48·40 (106)	8·82	7·33	10·13	22·64 (12)	7·25	5·89	8·79	56·63 (94)	9·13	7·49	10·32	**0·036**
HPV-33	7·76 (17)	7·54	6·27	10·57	1·89 (1)	9·33	n/a	n/a	9·64 (16)	7·19	5·43	11·27	0·540
HPV-45	39·73 (87)	6·62	5·96	9·44	5·66 (3)	5·95	5·31	9·35	50·60 (84)	6·63	5·97	9·51	0·340
HPV-58	47·03 (103)	6·00	5·66	6·26	20·75(11)	5·66	2·98	6·08	55·42 (92)	6·02	5·73	6·26	0·137

HPV, human papillomavirus; n/a, not applicable.
Absolute viral loads have been given in log10.
Values shown in bold denote statistically significant ones ($P < 0.05$).
* The Mann–Whitney test was used for comparing single and multiple infection for all HPV types.

Table 3. *Active variables' contributions, cosine squared, test values, and the corresponding type's viral load correlation coefficient*

Variable	Libel modality	Contribution		Square cosine		Test value		Viral load correlation coefficient	
		Axis 1	Axis 2	Axis 1	Axis 2	Axis 1	Axis 2	Axis 1	Axis 2
HPV-16	No	**13·83**	3·24	**0·45**	0·07	**−9·87**	4·02	0·65	−0·25
	Yes	**13·96**	3·27	**0·45**	0·07	**9·87**	−4·02		
HPV-18	No	1·69	4·27	0·05	0·08	−3·14	−4·21	0·17	0·19
	Yes	1·13	2·87	0·05	0·08	3·14	4·21		
HPV-31	No	7·77	**10·90**	0·26	0·26	7·50	**7·48**	−0·50	−0·43
	Yes	8·28	**11·62**	0·26	0·26	−7·50	**−7·48**		
HPV-33	No	0·30	2·72	0·06	**0·40**	−3·70	**9·32**	0·19	−0·55
	Yes	3·60	**32·29**	0·06	**0·40**	3·70	**−9·32**		
HPV-45	No	**14·79**	2·41	**0·60**	0·07	**−11·42**	3·88	0·65	−0·33
	Yes	**22·43**	3·66	**0·60**	0·07	**11·42**	−3·88		
HPV-58	No	5·75	**10·70**	0·20	**0·26**	6·55	**7·51**	−0·39	−0·50
	Yes	6·48	**12·05**	0·20	**0·26**	−6·55	**−7·51**		

HPV, human papillomavirus.
The values in bold show the axis where HPV types were representative, except for HPV-18 (which had low representation on both axes). The test value's sign indicates the pole of the respective axis where the variable was grouped.

births, infection by more than three viral types and the corresponding viral loads. Bogotá, being infected by three or less viral types and using no type of contraceptive were the variables associated with absence of infection by these types (Table 4).

The second profile (species related) so identified was represented by grouping HPV-31, HPV-33 and HPV-58 (A9 species) on axis 2. The closeness between HPV-33 and HPV-58 denoted greater association between these two types. The variables associated with infection by these types involved Bogotá, linked healthcare regime, hormonal contraceptives, secondary level of education, nulliparity, coinfection by more than three viral types, ASCUS and corresponding types' viral loads. On the other hand, when there was no infection by these three types, the representative variables involved Girardot, background of 3–4 pregnancies, primary school educational level, subsidised healthcare regime and coinfection by three viral types or less (Table 4).

The two-dimensional graph (Fig. 1) shows how illustrative nominal (socio-demographic characteristics, risk factors and cytological findings), illustrative continuous (viral loads) and active variables (viral types) were associated regarding profile formation. This graph is like a Cartesian plane formed by two axes, each having

Table 4. *Test values for illustrative nominal variables*

Variable	n	Test value Axis 1	Test value Axis 2
Origin			
Girardot	141	**2·83**	**3·15**
Chaparral	5	1·53	−0·08
Bogotá	73	**−3·36**	**−3·18**
Educational level			
Illiterate	5	−0·10	−0·07
Primary	107	0·53	**2·69**
Secondary	88	−0·22	**−2·58**
Technical qualification	11	0·66	−0·69
Graduate	8	−1·52	0·42
Healthcare scheme affiliation			
Private	14	0·47	−1·11
Contributory	5	−0·48	0·20
Subsidised	139	1·05	**3·03**
Linked	61	−1·23	**−2·71**
Contraceptive method			
Barrier	10	0·09	−0·22
IUD	27	0·36	−0·28
Hormonal	23	0·13	**−2·70**
Surgery	70	1·76	1·34
Contraceptive, none	89	**−2·03**	0·69
Pregnancies			
Pregnancies, 0	5	−0·81	**−2·09**
Pregnancies, 1–2	82	0·27	−1·61
Pregnancies, 3–4	100	0·13	**2·40**
Pregnancies, >4	32	−0·22	−0·30
Births			
Births, 0	6	−0·66	**−2·57**
Births, 1–2	109	−1·98	−0·33
Births, 3–4	92	**2·45**	1·10
Births, >4	12	−0·48	0·19
Cytological findings			
Cytology, negative	205	1·29	0·94
Cytology, ASCUS	5	−1·06	**−2·49**
Cytology, LSIL	9	−0·80	0·71
Number of viral types			
⩽3 HPV types	177	**−2·39**	**7·95**
>3 HPV types	42	**2·39**	**−7·95**

IUD, intrauterine device; ASCUS, atypical squamous cells of undetermined significance; LSIL, low-grade squamous intraepithelial lesion; HPV, human papillomavirus.
Test values were obtained from analysing every category for each variable on each factorial axis (or groups of HPV types). Such values indicate the location of coordinates for each category on a two-dimensional plane (constructed from the axes shown in Fig. 1). Significant representation was ⩽−2 or ⩾2 as cut-off points (values in bold). Variables associated with a particular HPV group were identified according to the test value sign (negative or positive).

a positive and negative pole. Each HPV type could be loaded onto each axis (either positively or negatively). This reflected the data given in Table 3; this showed

the cosine squared indicating representation, the contribution or association of each viral type on each axis and test value sign showing the position on a determined quadrant. These values (such as coordinates on a Cartesian plane) indicated the position and association of a variable within a profile. An arrow represented each type's viral load whose direction towards a determined quadrant indicated the association with a particular grouping and its length revealed the contribution within the profile. Regarding sociodemographic variables, risk factors and cytological findings (empty boxes), the proximity between boxes and the distance to viral types indicated the strength of their association.

DISCUSSION

MCA led to identifying and evaluating the profiles configured by the dependency relationships between HPV infection and other variables. Two grouping patterns were identified for the six most relevant HR-HPV types in the Colombian population. It was seen that two high-risk viral types (HPV-16 and −45) from different species (A9 and A7, respectively) were grouped and three types from the same species (HPV-31, −33 and −58). It was seen that HPV-16 was not grouped with types from the same species (HPV-31, −33 and −58) which could be related to competition between types; likewise, HPV-18, in spite of being the most frequently occurring viral type in both single and multiple infections, did not have a tendency to become grouped with either of these two profiles.

Previous studies have found negative associations for infection by another viral type when HPV-16 is present, as well as positive associations for other combinations [16]. The present study found two grouping profiles for HR-HPV types, each type being representative for the corresponding profile (except HPV-18). This cannot be used for proving a biological dependence between viral types; however, it might suggest that infection by some types could depend on the existence of one or more viral types and the interaction between coexisting types [17].

It has been reported that some type of competition between HPV types for the ecological niche is probable in natural coinfection [18, 19]; however, different viral types could coexist in the same cell population, or even in the same cell, whilst having interactions between them [17]. The viral types could compete in the cervical epithelium through factors necessary for
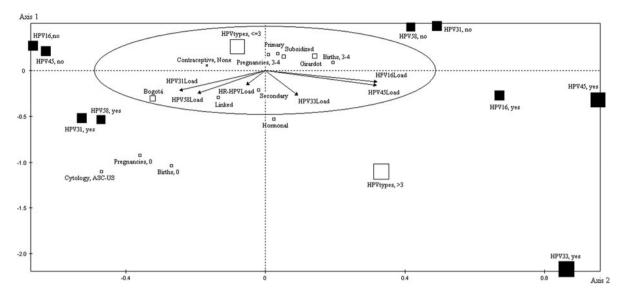
Fig. 1. Multiple correspondence analysis (MCA). The figure shows the association of the variables distributed in the four quadrants. Active variables are represented by filled-in squares, the illustrative nominal variables are shown by empty squares and arrows represent illustrative continuous variables. The size of the squares is proportional to each modality's contribution to the most representative factor. The arrow's length shows the correlation of viral load with each factor. The numbers on the axes represent test values, indicating the significance of representation of each modality's variable regarding each factor.

replication, transcription, translation and/or persistence [20]. Such competition could affect the replication of both genomes or just that of a single viral type which could be seen regarding differences in viral load [17]. However, a specie-specific type of association has been described where HPV-16 and HPV-18 viral loads became reduced in the presence of types from the same species, suggesting an interaction between coexisting types and the induction of a cross-reactive immune response [20]. Due to the homology regarding the genome's sequence and similar structure, HPV types could induce such responses and also favour a trend towards grouping [10, 21].

However, such differences between viral loads in coinfection could also represent viral integration to host genome [17]. HPV-16 had lower absolute viral load regarding coinfection compared with HPV-45. HPV-31 and HPV-58 viral loads were greater in coinfection compared with single infection, whilst HPV-33 had a lower load in multiple infection. These findings suggest an induced response against HPV-16 and HPV-33 or integration with host genome [17].

No competition between viral types from the same species was seen for the second profile (HPV-31, −33 and −58); this may have facilitated infection by the three types as cell regulators were available which promoted each viral type's replication, thereby leading to cooperativity between these types [17, 20]. Another

factor which could have facilitated the grouping of these viral types (from the same species) was infection temporality. It has been described that the probability of acquiring HPV infection increases if there is already infection by another high risk type [2, 3, 22]. HPV-31, HPV-33 and HPV-58 infection would thus have been subsequent, contrary to that observed for HPV-16 and HPV-31 (concurrent).

HPV-18 has frequently been reported in Colombia as being the second most prevalent type after HPV-16, contrary to that found in this study where it was reported most frequently. However, it is worth highlighting that women from just three cities in Colombia were included in the study and that such frequency changes depending on the region being evaluated according to the women's age and whether they have cervical lesions [11]. P&P programmes must thus pay special attention to such changes regarding prevalence due to HPV-18 being associated with 15% of CC cases [23].

One of the variables associated with clustering patterns was the origin. Differences between Girardot and Bogotá have been identified in relation to the prevalence and type-specific distribution of HPV, multiple infections (being higher for Girardot) and the time required for clearance (lower in Girardot when A9 species types were present) [11, 12]. However, so far there are no reports of previous studies about the

association of the origin of women with the grouping of HPV genotypes in coinfection. Factors related to sexual behaviour and cultural characteristics, which might be different between these cities, were not evaluated in this study but could modulate the associations found here.

Regarding parity, 3–4 births was associated with the presence of HPV-16 and HPV-45, this is consistent with the finding that multiparity have a higher risk of HPV infection [24]. This has been explained by the prolonged exposure to the contact with HPV of the transformation zone into the exocervix [25]. However, no history of births was found to be associated with HPV-31, HPV-33 and HPV-58 coinfection, this agreed with previous reports [11]. This could suggest that the risk of infection conferred by a history of births and pregnancies depends on viral genotype and probably sexual behaviour and contraceptive methods used by those women [25].

Our findings confirm what has previously been stated, that hormonal contraception is related to HPV infection [25]. As is known, HPV infection is sexually transmitted, it is likely that women that use a hormonal contraceptive method do not use barrier methods such as condoms, which would affect the transmission of the virus and therefore, facilitate the acquisition of various HPV types.

Regarding the healthcare scheme affiliation, the subsidised and linked regimes include people without capacity to pay, from low socio-economic strata [26]. The association between the linked regime and coinfection with HPV-31, HPV-33 and HPV-58, and also with secondary education, agreed with 85% of the HPV infection burden occurring in developing countries [27] as with studies that have shown that low-income women are more likely to have HPV [28]. Factors such a risky sexual behaviour [28] and possibly nutritional deficiencies [29] in low-income populations may make them more susceptible to HPV infection.

An association was found between the second profile and the presence of ASCUS in cytology. Three viral types were grouped in this profile, having absolute loads of 10^9, 10^7 and 10^6 viral copies (HPV-31, −33 and −58, respectively). It has been described that developing cervical lesions is directly proportional to the level of viral load [30–33]. It has also been found that the number of viral types has been significantly correlated with lesion severity [34]. The foregoing would explain ASCUS being found in this profile but not in the first one. However, no associations were found with other cytological findings,

probably due to the size of the samples from these groups. Coinfection prevalence in ASCUS patients has been reported as being 10%–40·7% [22, 35]. Spinillo et al. [36] suggested that the presence of coinfection is common in ASCUS and LSIL (low-grade SIL) women and is associated with a greater risk of invasive SIL and CC compared with single infection. The foregoing has implications for managing women having abnormal cytology.

As the present study was cross-sectional, most infection may have been prevalent but not incident; similarly, subsequent and concurrent infection could not be evaluated. However, the results from this study are consistent regarding some follow-up studies [2–4]. Moreover, no associated variables with multiple HPV infection were found when multivariate logistic regression or Poisson analyses were carried out. However, previous studies with a larger simple size in Colombian population have described factors such as women procedence (Andean region, Leticia), ethnicity (indigenous), birth or abortion antecedents, associated with multiple HPV infection [11, 37, 38]. Some of these variables configure the described profiles identified through the MCA.

This study has provided relevant information regarding multiple HPV infection, information that has not been evaluated or reported by other studies. As real-time PCR was used as the detection technique, infection having low viral load was probably identified which could not have been detected by a qualitative technique [3]. The present study dealt with the most relevant types for Colombia, meaning that information could be established as baseline for future evaluation of the impact of vaccination on the Colombian population. There is concern that viral types already covered (HPV-16, −18, −6 and −11) may become eliminated and positive selection may occur regarding viral types which are not covered by vaccines, increasing their frequency. Their interaction in natural infection must be taken into account as a tendency or grouping pattern in multiple infections, as has been shown [22, 39, 40]. Future research is needed regarding the mechanisms and other factors behind the relationships found between HPV genotypes.

The findings suggested that in coinfection, HR-HPV types tended to become grouped and the relationship between the grouping profiles was different regarding viral type. Biological characteristics such as the species to which they belonged, the similarity of their genomes and the number of viral copies, temporality regarding acquiring infection by different

HPV genotypes and competition between viral types could influence the configuration of different grouping patterns. Differently to previous studies, host characteristics were included allowing a better approach to such associations with the presence of one pattern or another. Other yet-to-be-evaluated factors may intervene in the interaction between different coexisting viral types, such as the women's susceptibility, which predisposes some females to HPV infection more than others, or such susceptibility could be type specific, involving synergism between viral types, immune system alterations and sexual companions' characteristics [4]. Within the configured profiles, there is an interaction with both, host and virus factors, which can be associated with specific patterns of grouping, reflecting the importance of evaluating both viral and host's characteristics when carrying out studies about infection by multiple HR-HPV types.

ACKNOWLEDGEMENTS

DECLARATION OF INTEREST

None.

ETHICAL STANDARDS

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

REFERENCES

1. **Bosch FX, et al.** The causal relation between human papillomavirus and cervical cancer. *Journal of Clinical Pathology* 2002; **55**: 244–265.

2. **Liaw KL, et al.** A prospective study of human papillomavirus (HPV) type 16 DNA detection by polymerase chain reaction and its association with acquisition and persistence of other HPV types. *Journal of Infectious Diseases* 2001; **183**: 8–15.

3. **Mendez F, et al.** Cervical coinfection with human papillomavirus (HPV) types and possible implications for the prevention of cervical cancer by HPV vaccines. *Journal of Infectious Diseases* 2005; **192**: 1158–1165.

4. **Thomas KK, et al.** Concurrent and sequential acquisition of different genital human papillomavirus types. *Journal of Infectious Diseases* 2000; **182**: 1097–1102.

5. **Chaturvedi AK, et al.** Human papillomavirus infection with multiple types: pattern of coinfection and risk of cervical disease. *Journal of Infectious Diseases* 2011; **203**: 910–920.

6. **Cuschieri KS, et al.** Multiple high risk HPV infections are common in cervical neoplasia and young women in a cervical screening population. *Journal of Clinical Pathology* 2004; **57**: 68–72.

7. **Garcia DA, et al.** Highly sensitive detection and genotyping of HPV by PCR multiplex and Luminex technology in a cohort of Colombian women with abnormal cytology. *Open Virology Journal* 2011; **5**: 70–79.

8. **Chaturvedi AK, et al.** Prevalence and clustering patterns of human papillomavirus genotypes in multiple infections. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 2005; **14**: 2439–2445.

9. **Trottier H, et al.** Human papillomavirus infections with multiple types and risk of cervical neoplasia. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 2006; **15**: 1274–1280.

10. **Vaccarella S, et al.** Clustering of multiple human papillomavirus infections in women from a population-based study in Guanacaste, Costa Rica. *Journal of Infectious Diseases* 2011; **204**: 385–390.

11. **Camargo M, et al.** Frequency of human papillomavirus infection, coinfection, and association with different risk factors in Colombia. *Annals of Epidemiology* 2011; **21**: 204–213.

12. **Soto-De Leon SC, et al.** Persistence, clearance and reinfection regarding six high risk human papillomavirus types in Colombian women: a follow-up study. *BMC Infect Dis* 2014; **14**: 395.

13. **Carcopino X, et al.** Determination of HPV type 16 and 18 viral load in cervical smears of women referred to colposcopy. *Journal of Medical Virology* 2006; **78**: 1131–1140.

14. **Escofier B, Pagès J.** *Análisis factoriales simples y múltiples: objetivos, métodos e interpretación*. España: Universidad del País Vasco, 1992, p. 286.

15. **Lebart L, Morineau A, Piron M.** *Statistique Exploratoire Multidimensionnelle*. Paris: Dunod, 1995, p. 439.

16. **Mejlhede N, et al.** Multiple human papilloma virus types in cervical infections: competition or synergy? *APMIS: Acta Pathologica, Microbiologica, et Immunologica Scandinavica* 2010; **118**: 346–352.

17. **Dickson EL, et al.** Cervical cytology and multiple type HPV infection: a study of 8182 women ages 31–65. *Gynecologic Oncology* 2014; **133**: 405–408.

18. **Dickson EL, et al.** Multiple-type human papillomavirus (HPV) infections: a cross-sectional analysis of the prevalence of specific types in 309 000 women referred for HPV testing at the time of cervical cytology. *International Journal of Gynecological Cancer: Official Journal of the International Gynecological Cancer Society* 2013; **23**: 1295–1302.

19. **Murall CL, McCann KS, Bauch CT.** Revising ecological assumptions about Human papillomavirus interactions and type replacement. *Journal of Theoretical Biology* 2014; **350**: 98–109.

20. **Xi LF, et al.** Human papillomavirus types 16 and 18 DNA load in relation to coexistence of other types, particularly those in the same species. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 2009; **18**: 2507–2512.

21. **McCarthy C, Youde SJ, Man S.** Definition of an HPV18/45 cross-reactive human T-cell epitope after DNA immunisation of HLA-A2/KB transgenic mice. *International Journal of Cancer Journal International du Cancer* 2006; **118**: 2514–2521.

22. **Rousseau MC, et al.** Occurrence of cervical infection with multiple human papillomavirus types is associated with age and cytologic abnormalities. *Sexually Transmitted Diseases* 2003; **30**: 581–587.

23. **Clifford G, Franceschi S.** Members of the human papillomavirus type 18 family (alpha-7 species) share a common association with adenocarcinoma of the cervix. *International Journal of Cancer* 2008; **122**: 1684–1685.

24. **Zhang R, et al.** Risk factors for human papillomavirus infection in Shanghai suburbs: a population-based study with 10 000 women. *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology* 2013; **58**: 144–148.

25. **Castellsague X, Bosch FX, Munoz N.** Environmental co-factors in HPV carcinogenesis. *Virus Research* 2002; **89**: 191–199.

26. **Guerrero R, et al.** Sistema de salud de Colombia. *Salud Pública de México* 2011; **53**: 144–155.

27. **Ferlay J.** *Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 11*. Lyon, France: International Agency for Research on Cancer, 2013.

28. **Schluterman NH, et al.** Differences in patterns of high-risk human papillomavirus infection between urban and rural low-resource settings: cross-sectional findings from Mali. *BMC Women's Health* 2013; **13**: 1–9.

29. **Goodman MT, et al.** Hawaii cohort study of serum micronutrient concentrations and clearance of incident oncogenic human papillomavirus infection of the cervix. *Cancer Research* 2007; **67**: 5987–5996.

30. **Hernández-Hernández DM, et al.** Association between high-risk human papillomavirus DNA load and precursor lesions of cervical cancer in Mexican women. *Gynecologic Oncology* 2003; **90**: 310–317.

31. **Josefsson AM, et al.** Viral load of human papilloma virus 16 as a determinant for development of cervical carcinoma in situ: a nested case-control study. *Lancet* 2000; **355**: 2189–2193.

32. **Moberg M, et al.** High viral loads of human papillomavirus predict risk of invasive cervical carcinoma. *British Journal of Cancer* 2005; **92**: 891–894.

33. **Ylitalo N, et al.** Consistent high viral load of human papillomavirus 16 and risk of cervical carcinoma in situ: a nested case-control study. *Lancet* 2000; **355**: 2194–2198.

34. **Bello BD, et al.** Cervical infections by multiple human papillomavirus (HPV) genotypes: prevalence and impact on the risk of precancerous epithelial lesions. *Journal of Medical Virology* 2009; **81**: 703–712.

35. **Schmitt M, et al.** Viral load of high-risk human papillomaviruses as reliable clinical predictor for the presence of cervical lesions. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 2013; **22**: 406–414.

36. **Spinillo A, et al.** Multiple human papillomavirus infection and high grade cervical intraepithelial neoplasia among women with cytological diagnosis of atypical squamous cells of undetermined significance or low grade squamous intraepithelial lesions. *Gynecologic Oncology* 2009; **113**: 115–119.

37. **Soto-De Leon S, et al.** Distribution patterns of infection with multiple types of human papillomaviruses and their association with risk factors. *PloS ONE* 2011; **6**: e14705.

38. **Del Rio-Ospina L, et al.** The prevalence of high-risk HPV types and factors determining infection in female Colombian adolescents. *PloS ONE* 2016; **11**: e0166502.

39. **Rousseau MC, et al.** Cervical coinfection with human papillomavirus (HPV) types as a predictor of acquisition and persistence of HPV infection. *Journal of Infectious Diseases* 2001; **184**: 1508–1517.

40. **Tota JE, et al.** Epidemiologic approaches to evaluating the potential for human papillomavirus type replacement postvaccination. *American Journal of Epidemiology* 2013; **178**: 625–634.