INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI®
800-521-0600

RISK FACTORS FOR INCIDENT CERVICAL HUMAN PAPILLOMAVIRUS INFECTION IN WOMEN IN A HIGH-RISK AREA FOR CERVICAL CANCER

Marie-Claude Rousseau

Department of Epidemiology & Biostatistics
McGill University, Montreal

February 1998

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfilment of the requirements for a Master of Science degree

ABSTRACT

Human papillomavirus (HPV) is the sexually-transmitted etiologic agent of cervical cancer. Despite screening programs, cervical cancer remains too common, particularly in developing countries. Various correlates of prevalent infections have been identified. However, the determinants of incident infections have never been studied.

Data were collected during a prospective cohort study conducted in Brazil. Incidence density rates of infection were calculated and determinants of incident infection were identified using Cox regression models. Analyses were done for HPV types classified into low-risk and high-risk depending on their association with cervical neoplasia.

The incidence density rates were 9.3 and 7.6 per 1000 women-months respectively for low-risk and high-risk HPV infection. Independent positive associations were found between the time of first occurrence of low-risk infection and age, number of sexual partners in the past 5 years, education level and use of non-commercial hygienic absorbents. The first occurrence of high-risk infection was independently predicted by age, age at first sexual intercourse, condom use (negative associations) and by the number of sexual partners in the past year (positive association). Elucidation of the dynamics of infection is a first step towards implementation of public health programs for reducing the risk of cervical cancer.

RÉSUMÉ

Le virus du papillome humain (HPV) est identifié comme l'agent causal sexuellement transmissible du cancer du col utérin. Malgré les programmes de dépistage, ce cancer demeure trop fréquent, particulièrement dans les pays non-industrialisés. De nombreux facteurs de risque des infections prévalentes par le HPV ont été identifiés. Cependant, aucune étude n'a porté sur l'identification des déterminants des infections incidentes par ce virus.

Les données ont été recueillies lors d'une étude prospective de cohorte au Brésil. La densité d'incidence d'infection a été calculée et les déterminants des infections incidentes ont été identifiés à l'aide de modèles de régression de Cox. Ces analyses ont été effectuées pour deux grandes classes de HPV définies sur la base de leur association avec le cancer du col utérin: les HPV oncogéniques et non-oncogéniques.

Les densités d'incidence obtenues étaient respectivement de 9,3 et de 7,6 par 1000 personnes-mois pour les infections par des HPV oncogéniques et non-oncogéniques. Des associations positives indépendantes ont été obtenues entre le temps de la première infection par des HPV non-oncogéniques et l'âge, le nombre de partenaires sexuels durant les 5 dernières années, le niveau de scolarité, ainsi que l'utilisation de serviettes hygiéniques non commerciales. Le risque de première occurrence d'une infection par des HPV oncogéniques a été prédit de façon indépendante par l'âge, l'âge lors de la première relation sexuelle, l'utilisation du condom (associations négatives) ainsi que par le nombre de partenaires sexuels durant la dernière année (association positive). L'élucidation de la dynamique des infections permettra de franchir une première étape vers l'établissement de programmes de santé publique visant la réduction du risque de cancer du col utérin.

STATEMENT OF ORIGINALITY

The research described in this thesis consists in original work. To our knowledge, incidence density rates of HPV infection have been published only from one previous study. However, these rates were not available upon stratification by oncogenic risk and by age. The analyses presented here can therefore shed new light on the specific characteristics of these infections. More importantly this project is, to our knowledge, the first one to aim at identifying the determinants of first occurrence of genital HPV infection with low-risk and high-risk types.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (CONTINUED)

LIST OF FIGURES

LIST OF APPENDICES

INTRODUCTION

From a worldwide perspective, cancer of the cervix is the second most common cancer in women, representing 11.6% of female cancers compared to 19.1% for breast cancer. In industrialized countries where formal screening programs are available, cervical cancer is tenth in overall importance. Almost 80% of the incident cervical cancer cases occur in developing countries, where this cancer is the most frequent one (Parkin et al.,1993).

Early epidemiological studies on risk factors for cervical cancer showed a strong and consistent association between markers of sexual activity and cervical cancer or its precursor lesions (Brinton and Fraumeni,1986; Brinton et al.,1987; LaVecchia et al.,1986; Brinton et al.,1989; Brock et al.,1989; Slattery et al.,1989). This main risk factor led epidemiologists to suspect a sexually-transmitted etiologic agent. The human papillomavirus (HPV) was the most carefully studied agent in relation to cervical cancer. Some initial studies did not show correlations between HPV infection and sexual activity (Villa and Franco,1989; Kjaer et al.,1990) or between HPV infection and cervical cancer (Reeves et al.,1989); however, this was later found to be due to misclassification of HPV status (Franco,1991; Franco,1992). Epidemiologic studies using more sensitive and specific detection methods revealed the association between HPV and risk of cervical cancer (Munoz and Bosch,1992; Koutsky et al.,1992). HPV types are often classified based on the degree of their association with invasive cervical cancer. The most commonly used definition separates high-risk (types 16, 18, 31, 33, 45, 51, 52, 56, 58) from low-risk types (types 6/11, 40, 42, 53, 54, 55, 57, 59, isolates awaiting taxonomic entry, unidentified types) (Bauer et al.,1993). Nowadays, there is a vast body of knowledge, both epidemiological and biological, supporting the causal role of HPV in cervical neoplasia. The evidence implicates HPV as the primary cause of cervical cancer, but also indicates that HPV infection is not sufficient for the development of this multifactorial disease.

HPV is responsible for a ubiquitous sexually-transmitted infection, with a prevalence that can reach 45% among cytologically normal women when polymerase chain reaction (PCR)-based methods of detection are used (IARC Working Group, 1995). However, the vast majority of these infections does not result in cellular abnormalities.

Most studies examining HPV infection are based on cross-sectional designs, giving a static picture of a phenomenon that is in fact quite dynamic. Infections caused by HPV are mainly transient (Evander et al., 1995) and it has been suggested that persistent infections might be the ones from which cervical neoplasia arises (Hildesheim et al., 1994). Before studying persistent infections, which requires a substantial accumulation of follow-up due to its rarity, a more frequent event such as incident infection could be studied, and would allow a first insight into the dynamic nature of HPV. By taking into account the temporal relationship and dynamic nature of infection, the study of incidence could help elucidate the modes of transmission and acquisition of HPV. Such information will be most useful in the planning of public health programs aiming at primary prevention of cervical cancer (Franco, 1997).

The objectives of this research project are therefore to: 1) calculate the incidence-density rates for overall, low-risk, and high-risk cervical HPV infections using person-time denominators, and; 2) identify the determinants of incident infections for low-risk and high-risk HPV types.

1 Cervical cancer and its link to human papillomaviruses

1.1 Epidemiology of cervical cancer

From a worldwide perspective, cervical cancer ranks fifth in terms of frequency. Approximately 437,000 new cases are diagnosed each year. Cervical cancer is the second-most common cancer in women representing 11.6% of all incident cancer cases, just after breast cancer (19.1%). In developed countries, cancer of the cervix ranks tenth in overall importance, due to the availability of general screening programs and low fertility rates. However, in developing countries, it is the most frequent cancer. Almost 80% of all cervical cancer cases occur in developing countries, with an estimated 344,000 cases per year diagnosed in those areas not systematically providing screening programs (Parkin et al., 1993). The estimated age-standardized incidence rates of cervical cancer for 1985 vary from 7.6 per 100,000 in Western Asia to 46.8 per 100,000 in Southern Africa, with rates being generally less than 15 per 100,000 (Parkin et al., 1993). Incidence and death rates vary also according to ethnicity: rates for African American, Hispanic, and Native women can be as high as 2 to 3 times those of white women (Gusberg and Runowicz, 1991). The higher incidence of cervical cancer in some ethnic groups is very likely confounded by socioeconomic status and related issues such as lifestyle, nutrition, attitude towards health care, and access to screening.

In past decades morbidity and mortality due to cervical cancer have declined, mostly in industrialized countries. Important contributing factors were undoubtedly the implementation of general cytological screening programs and the decline in fertility that occurred after the mid 1950s. However, cervical cancer still generates severe consequences. The overall 5-year survival rate for women with cervical cancer is 65%, 10% lower than the one for breast cancer (Gusberg and Runowicz, 1991). Furthermore, data from Statistics Canada show that on average, new cases of invasive cervical cancer are approximately 10 years younger than women with breast cancer (50 compared to 60 years of age) and associated death occurs also

at a younger age. Cervical cancer affects multiparous and relatively young women, resulting in important losses for society (Hagey et al.,1995).

1.2 Risk factors for cervical cancer

The most consistently reported risk factor for cervical cancer is undoubtedly sexual behaviour. Whether from case series, case-control or cohort studies, the importance of diverse markers of sexual behaviour has been outstanding. One of the first reports of such an association with sexual activity is attributed to Dr. Rigoni-Stern, who, in 1842 described the difference in proportions of death due to uterine cancer in nuns and women with varying marital status. A much smaller proportion of death due to uterine cancer, comprising at that time cancer of the cervix and body of the uterus, was seen among nuns than among married women (original Italian article translated by De Stavola,1987). Marital status was one of the first proxies of sexual activity in early observational studies of cervical cancer. Young age at marriage and marriage itself were also identified as factors related to cancer of the cervix (Wynder et al.,1954). Ecological studies also found correlations between morbidity (Franco et al.,1988) and mortality (Li et al.,1982) of penile and cervical cancer, indirectly suggesting the association with sexual activity. Following these findings, many epidemiological studies conducted in various geographical areas and on different populations focussed on a certain number of markers of the multidimensional and complex factor that is sexual behaviour. Two markers have been extensively used and studied: the number of sexual partners and the age at first sexual intercourse. The number of sexual partners is undoubtedly an important factor for cervical cancer, with the risk of cancer being close to 3 times higher in women with 10 or more lifetime sexual partners compared to women with 0 or 1 partner (Brinton et al.,1987; Peters et al.,1986). Numerous case-control studies have addressed this issue, with varying results due to the wide range of populations, selection criteria, and control of confounding made in the analysis. However, number of sexual partners seemed consistently associated with cervical cancer, the odds ratios (ORs) ranging from 1.5 to 10.0 for women with 5, 6, 7, even 10 or more lifetime partners compared to women with very few partners (Brinton et al.,1987; Brock et al.,1989; Slattery et al.,1989; Herrero et al.,1990).

In general, the risk of invasive or pre-invasive cancer has been described as twice as high in women who initiated sexual life earlier (16 years or less in general) compared to women who had their first sexual intercourse in their twenties. Most of these studies show an effect even after adjustment for number of sexual partners, a factor frequently correlated with age at first intercourse (Brinton et al.,1987; Brock et al.,1989; Herrero et al.,1990; LaVecchia et al.,1986). Based on this independent effect, it was hypothesized that the cervix might be more vulnerable towards carcinogenic agents during adolescence when metaplasia is more likely to occur (Brinton,1992).

A few studies have addressed the effect of the male partner's sexual behaviour, comparing partners of women with and without cervical cancer. Partners of the former consistently reported more sexual partners (Brinton et al.,1989) and contacts with prostitutes than partners of women free of cervical disease (Kjaer et al.,1991). Partners of women with cervical cancer were also less likely to have used condoms regularly (Kjaer et al.,1991). Case-control studies conducted in Spain and Colombia, respectively low-risk and high-risk areas for cervical cancer, also addressed the sexual behaviour of male partners. In Spain, the risk of cervical neoplasia was strongly associated with the number of other partners (OR = 11.0; 95%CI:3.0-30.0 for $\geq$ 21 women vs 1), and to the number of prostitutes (OR = 8.0; 95%CI:2.9-22.2; for $\geq$10 women vs 0) (Bosch et al.,1996). In Colombia, no significant difference was found between the sexual behaviour of husbands of cases and controls. The authors explained that the relatively high proportion of men having extramarital relations and visiting prostitutes resulted in insufficient statistical power to detect small differences between the male partners of case and control women (Munoz et al.,1996).

Some of the early epidemiological studies suggested an association between cigarette smoking and cancer of the cervix (Naguib et al.,1966; Thomas,1973). However, since smoking is often correlated to sexual behaviour, it has been proposed that these early results were confounded by the effect of sexual activity. Subsequent studies in which the investigators had the necessary information to control for sexual behaviour by adjusting for age at first intercourse

and number of sexual partners, showed an independent effect of cigarette smoking (Brinton and Fraumeni,1986; LaVecchia et al.,1986; Slattery et al.,1989). These findings were substantiated by the presence of high levels of nicotine metabolites in the cervical mucus of smokers, therefore suggesting a possible direct carcinogenic or initiating effect on the epithelium (Schiffman et al.,1987). The immunosuppressive effect of smoking has been suggested as another mechanism of action (Barton et al.,1988).

An independent effect of the number of live births has been demonstrated in two large case-control studies, showing a linear trend of increased risk for cervical cancer as the number of live births increased (Brinton et al.,1987; Brinton et al.,1989).

Controversial results have been obtained concerning the association of oral contraceptive (OC) use with cervical cancer. While early studies found no excess risk of cancer among OC users, some of the more recent ones indicated elevated risks particularly among long-term users (Beral et al.,1988; Brinton and Fraumeni,1986; Brinton,1992). Problems that arise in studying OC use are the confounding effects of sexual activity and active screening leading to a very likely detection bias. Infrequent Pap smear screening is associated with a higher risk of cervical cancer.

Studies of dietary intake in relation to cervical cancer have mainly suggested a protective effect of vitamin C and beta-carotene (Herrero et al.,1991; VanEenwyk et al.,1992), except from a large case-control study conducted in 5 American areas that showed no such associations (Ziegler et al.,1991).

Immunosuppression, whether clinically-induced among organ transplant patients or resulting from HIV infection is a determinant of development of neoplastic precursors and cancer of the cervix. Because of the important morbidity resulting from cervical cancer among HIV-positive women, it is now recognized as an AIDS-defining illness by the Centres for Disease Control (Palefsky and Holly,1995).

1.3 Human papillomavirus in the etiology of cervical cancer

The strong and consistent association between cervical cancer and sexual activity led epidemiologists to suspect a sexually-transmitted agent. The first viral etiologic agent suspected in cervical cancer was the herpes simplex virus (HSV). The evidence in favour of this hypothesis came mainly from seroepidemiological surveys. Most studies found elevated levels of HSV type 2 (HSV-2) antibodies in the serum from cervical cancer patients when compared with normal subjects (Graham et al., 1982). However, early studies failed to detect viral DNA in the host cells with hybridization techniques (zur Hausen et al., 1974). Due to the improvement of DNA hybridization methods and to more specific HSV-2 molecular probes, herpes virus DNA could be detected in 10% to 40% of genital tumours, while viral DNA was absent from the tissue of similar organs from healthy controls subjects (Prakash et al., 1985; DiLuca et al., 1987). HSV possessed several properties of carcinogens, but some arguments were against an oncogenic role for this family of viruses. These arguments included a loss of viral DNA sequences from transformed cells after several passages in culture and a loss of viral antigenic expression in the host cells after some time in culture (DiLuca et al., 1987; Vonka et al., 1987). Those differences from the classic behaviour of oncogenic viruses and the fact that HSV-2 DNA was not consistently detected in tumour tissues led investigators to think that its role was more similar to that of a chemical carcinogen: limited to the initiation stage of malignancy (Skinner, 1976; reviewed in Franco, 1991).

In the mid-1970s, zur Hausen first formulated the hypothesis that cervical cancer might be caused by HPV infection (zur Hausen, 1976). In the late 1970s and 1980s, several genotypes of HPV were identified (Gissmann et al., 1977; Orth et al., 1978). The viruses responsible for most genital warts, HPV 6 and HPV 11, were also described by Gismann and zur Hausen in 1980 (reviewed in zur Hausen, 1994). The most important highly oncogenic HPV types were discovered in cervical tumour-related cells in 1983 and 1984. HPV 16 was identified from a cervical carcinoma and was subsequently found in tumour biopsies from various geographical areas (Dürst et al., 1983). HPV 18 was identified in HeLa cells, a cultured cell line derived from cervical tumour cells (Boshart et al., 1984). It soon became evident that the majority of

cervical tumours harboured HPV DNA sequences (Resnick et al.,1990; zur Hausen,1989; Schiffman et al.,1993).

Experimental studies showed evidence that the malignant cell changes were closely related to the expression of specific genes of the HPV genome, E6 and E7 (Crook et al.,1989; von Knebel Doeberitz et al.,1992). In transgenic mice, the expression of E6 or E7 alone resulted in development of condyloma, intraepithelial neoplasia and invasive squamous cell cancer on the skin of the mice (Lambert et al.,1993). The affinity of E6 and E7 gene products for tumour-suppressor proteins and resulting disruption of the normal cell cycle was experimentally demonstrated (Werness et al.,1990; Scheffner et al.,1990; Dyson et al.,1989; Münger et al.,1989). Several epidemiological studies identified high-risk HPV infections as the major risk factor for neoplasia of the cervix, with the association being strong and consistent (Munoz and Bosch,1992). Furthermore, in light of the epidemiological and biological data on the association between HPV infection and cervical cancer, a consensus panel convened by the International Agency for Research on Cancer (IARC) of the World Health Organization (WHO) concluded that there was sufficient evidence to identify HPV types 16 and 18 as *carcinogenic to humans*. HPV Types 31 and 33 have been identified as *probably carcinogenic*, and types other than 16, 18, 31, and 33 as possibly carcinogenic except HPV 6 and 11 (IARC Working Group,1995).

1.4 Screening for cervical cancer and its precursors

The Papanicolaou test or Pap smear, has proven to be an important tool in decreasing both morbidity and mortality from cervical cancer. This screening procedure consists in qualitative analysis of cells scraped from the transformation zone of the cervix. The transformation zone or squamo-columnar junction is the region of the cervix where the internal secreting columnar epithelium and the external stratified squamous epithelium meet. The junction between the 2 types of epithelium is located at the cervical os in young women. With age, the junction is replaced by metaplastic squamous epithelium and its location changes. From the ectocervix in young women, it withdraws further into the cervical canal, in the endocervix. This area of

the cervix is where most of the neoplastic lesions develop. For the Pap cytology, cells from this region are scraped, smeared and fixed on a microscope slide, for cytopathological analysis. Slides are then examined for cellular morphological abnormalities. The Pap test is widely available in industrialized countries which have a well organized health care system. To this day, many developing countries lack implementation of this simple though efficacious screening procedure (Richart, 1995). Incidence rates and morbidity due to cervical cancer have decreased markedly since the implementation of screening programs. In Canada, for the 25-year period between 1969 and 1993, the age-standardized[1] incidence rate of cervical cancer decreased from 21.6 per 100,000 to 8.9 per 100,000 individuals. For the same period, the Canadian age-standardized mortality rate also decreased from 7.4 per 100,000 to 2.6 per 100,000 individuals (McLaughlin et al., 1996).

The addition of HPV testing to current cytological screening procedures is subject to debate. The high prevalence of HPV DNA found in cytologically normal women, which will not necessarily result in abnormal lesions, has been cited as one of the arguments against the general use of HPV testing. Such a procedure would result in women being alarmed by a positive test that would very likely have a low positive predictive value for development of cytological abnormalities (Nuovo and Nuovo, 1991). However, the same authors view HPV testing as potentially useful for clinical management or identification of colposcopically- or histologically-identified abnormalities (Nuovo and Nuovo, 1991). Proponents of HPV testing argue that the possibility to assess the clinical relevance of mild lesions and the possibility to increase both sensitivity and specificity of screening would exist if such a laboratory method was used as an adjunct to cytological screening (Reid and Lorincz, 1991).

Several groups of experts have addressed the question of HPV testing as an adjunct to existing screening programs for cervical cancer. At a National Workshop on Screening for Cancer of the Cervix held in Canada in 1989, the participants came to the conclusion that the

---

[1]Standardized to the age distribution of the 1991 Canadian population

evidence did not support the implementation of laboratory diagnostic assays for HPV infection as a necessary addition to the current national screening program (Miller et al.,1991). During a workshop held in Brussels in November 1991, the IARC participants concluded, in view of the debate and divergence of opinions, that further research was needed to determine the usefulness of HPV testing in cervical cancer prevention. It was also decided that carefully planned and well-controlled HPV screening trials should be conducted before taking any further action (Franco,1992). Again in 1995, the Canadian Task Force on the Periodic Health Examination concluded that there was "fair evidence" not to include HPV testing in periodic health examinations. However, they emphasized the need for further research on HPV infection, mainly focussing on natural history and epidemiology, immunologic therapies, and screening (Johnson,1995).

1.5 Classification of precursor lesions

Results from cytological screening are expressed according to a formal classification of lesions. Such classification was modified over time and its evolution is a good reflection of the progressive state of knowledge on cervical cancer (Wright and Kurman,1996).

The term dysplasia was introduced by Reagan to describe the range of cellular morphology between a normal epithelium and carcinoma *in situ* (CIS). The dysplasia was categorized as mild, moderate, or severe "based on the degree of atypia and relative thickness of the abnormal cellular changes in the epithelium" (Reagan et al., 1953).

Following the advent of cytological screening, it was demonstrated that the incidence of dysplasia was much higher than that of invasive cervical cancer. A small proportion of women with dysplasia were also shown to experience progression to invasive cancer, but the majority of dysplasia seemed to regress spontaneously (Stern and Neely,1964). From this point on, dysplasia and CIS were viewed as two separate entities, the latter being very likely to progress to invasive cervical cancer.

Technological tools permitting, the 1960s saw the apparition of results from electron microscopy and tissue culture studies. The results suggested that dysplastic cells and cells from CIS were qualitatively very similar (Richart,1966; Shingleton et al.,1968). The epidemiological basis for this view of all grades of dysplasia as a same disease, with each a risk of progression to cervical cancer, was based on a study of 557 women which showed very little regression (only 6%) and a progression of 50% of the mild dysplastic lesions over follow-up that lasted 36 months on average (Richart and Barron,1969). However, the main selection criteria, 3 previous Pap smears showing mild dysplasia indicates a selection bias towards persistent and more "serious" mild dysplasia cases. The high progression and low regression rates of mild dysplasia would be refuted later by various studies (Nasiell et al.,1986; Heinzl et al.,1982). Nevertheless, Richart proposed that the different manifestations of HPV infection were different stages of the same disease process leading to squamous cell carcinoma (Richart,1969). The cervical intraepithelial neoplasia (CIN) terminology, based on this concept, classified lesions according to their size, location, and degree of dysplasia: mild cellular dysplasia was CIN 1, moderate cellular dysplasia was CIN 2, severe cellular dysplasia and carcinoma *in situ* (CIS) were CIN 3. This classification, referred to as the WHO classification, had the advantage of grouping CIS in the same category as severe dysplasia from which it was very difficult to differentiate clinically.

As the knowledge base on cervical cancer and HPV grew, it became clear that the reasoning on which the previous classification had been built was wrong. Two different disease entities were at play. Mild dysplastic lesions were the expression of a productive viral infection, whereas moderate and severe dysplastic lesions were true precursors of invasive cervical cancer (Wright et al.,1994). A new classification for cytologic diagnoses was agreed upon in the late 1980s (Solomon,1989). The Bethesda System was more up-to-date in terms of implication of HPV in cervical carcinogenesis, and was designed to lead to less misclassification of abnormalities. The terms CIN 1, 2, and 3 were replaced by two terms: low-grade squamous intraepithelial lesion (LSIL) and high-grade squamous intraepithelial lesion (HSIL). LSILs included those lesions that resulted from productive HPV infections and

that often showed specific cytological features such as: nuclear enlargement, hyperchromatic and angulated nuclei, multinucleation and cytoplasmic cavitation. Flat condyloma were included in LSIL, corresponding also to the result of a productive viral infection. HSIL regrouped CIN 2 and 3, which were both often correlated with the presence of intermediate- or high-risk HPV infection, rarely showed distinctive cytological signs of productive HPV infection, and presented a lower viral DNA copy number per cell (Wright and Kurman, 1996). Table 1 shows the correspondence between the 2 classifications:

Table 1: Cytomorphological classifications

Cytomorphology	*WHO classification*	*The Bethesda System*
normal	normal	WNL[1]
inflammation	normal	ASCUS/AGUS[2]
mild dysplasia	CIN 1	LSIL
moderate dysplasia	CIN 2	HSIL
severe dysplasia	CIN 3	
carcinoma *in situ*		
invasive cervical cancer	invasive cervical cancer	invasive cervical cancer

1 Within Normal Limits
2 Atypical Squamous Cells of Undetermined Significance/Atypical Glandular cells of Undetermined Significance

1.6 Treatments

When abnormal lesions are found by cytology, treatments for non-invasive cervical cancer precursors, mostly HSIL, are normally undertaken after biopsy-confirmation. Two methods used for biopsy are also undertaken clinically as excisional treatments. Conization, consisting of a cone-shaped biopsy of the epithelium can be done for cones of varying sizes according to the extent of the lesion. Electrosurgical loop excision is performed with the aid of colposcopy instruments. Small loop electrodes are used to cut and cauterize the tissue meant

for excision. Treatment can also be termed ablative and would then consist of the destruction rather than the excision of the lesion. Cryosurgery procedures imply destruction of the targeted cells by freezing below -22°C with a cryoprobe. Healing normally occurs within 12 weeks. An alternative ablation method is by CO_2 laser which causes destruction of affected cells due to the high level of energy transferred to tissues and immediate rise in heat (Wright et al.,1994).

2 Human papillomaviruses

Papillomaviruses have been extensively studied by molecular biologists; the ultimate purpose being to understand the process by which they induce cancer. The progress in the molecular biology of papillomaviruses has been hampered by the inability to harvest virus particles in tissue culture, most likely due to the viruses' affinity for well-differentiated cells, a characteristic lost in cell culture.

Papillomaviruses are small, non-enveloped, icosahedral DNA viruses with a marked tropism for squamous epithelial cells. They induce warts in a variety of higher vertebrates, including humans. Papillomaviruses have been characterized from various species including humans, cattle, rabbits, horses, dogs, and mice (Howley,1996).

2.1 Taxonomy

Taxonomically, papillomaviruses are classified in the *Papovaviridae* family along with polyomaviruses (Murphy et al.,1995). Viruses in this family all have the following characteristics: small size, absence of a viral envelope, icosahedral capsid, circular double-stranded DNA genome, and replication that takes place in the nucleus. The *Papovaviridae* family is further divided into 2 subfamilies, *Papillomavirinae* and *Polyomavirinae*, based on fundamental differences in genomic organization and biology (Howley,1996).

2.2 Classification

More than 70 distinct HPVs have been identified (De Villiers,1994), and evidence suggests the existence of additional novel HPV types (Berkhout et al.,1995; Bernard et al.,1994; Shamanin et al.,1994). HPV types are classified by genotype rather than by serotype. Types were originally defined by DNA:DNA hybridization with known HPVs, a genome showing 50% or less homology being identified as novel. Subtypes were identified when novel sequences of HPV DNA resulted in more than 50%, but less than 100% homology with a known type (Coggin and zur Hausen,1979). Limitations of this classification method were due to inconsistent correlations between genomic homology as shown by hybridization and actual nucleotide sequence homology (Chan et al.,1995). Such problems were resolved by a new set of classification criteria instituted by the Papillomavirus Nomenclature Committee, based on nucleotide sequence homology for 3 HPV genes: E6, E7, and L1. A novel HPV type is therefore identified when nucleotide divergence with any previously known type is more than 10% for the sequences of E6, E7, and L1 open reading frames (ORFs) (Delius and Hofmann,1994). When divergence is between 2% and 10%, a subtype is identified. An HPV for which divergence is 2% or less, but that does not exhibit perfect homology with a known type, is classified as a molecular variant of this particular type (Bernard et al.,1994).

Human papillomavirus types can be divided into 2 categories based on virus affinity for: 1) "moist and sparsely keratinised" mucosal epithelia, and; 2) "dry and fully keratinised skin" (von Krogh,1991). Belonging to the first category, genital HPVs can be referred to as low-, intermediate-, or high-risk types, based on their epidemiological associations with precursors or neoplastic lesions of the cervix. The low-risk group includes types 6, 11, 42, 43, and 44. These types are common in low-grade squamous intraepithelial lesions (LSIL), less common in high-grade squamous intraepithelial lesions (HSIL), and virtually absent in invasive cancer. The intermediate-risk group includes HPV types 31, 33, 35, 51, and 52. HPVs from this group are commonly detected in association with HSIL, but less so with LSIL or invasive cancer. The high-risk group includes HPV types 16, 18, 45, and 56 which are strongly associated with invasive cancer. They present differences in their association with HSIL, types

18, 45, and 56 being uncommon in such lesions and type 16 being equally predictive of HSIL and invasive cancer (Lorincz et al.,1992). Another definition of HPV also based on associations with invasive cancer is the most commonly used one. It defines HPV as low-risk or high-risk types. Types 6/11, 40, 42, 53, 54, 55, 57, 59, isolates awaiting taxonomic entry, and unidentified types are classified as low-risk ones, whereas the high-risk category includes types 16, 18, 31, 33, 45, 51, 52, 56, and 58 (Bauer et al.,1993).

2.3 Virion structure

Viral particles are approximately 55 nm in diameter. The genome is approximately 8,000 base pairs and the capsid is composed of 72 capsomeres. The capsid is made of 2 structural proteins: L1 and L2. L1 represents approximately 80% of the total capsid proteins, and is structurally very important. The morphology of virus-like particles (VLPs) when produced from L1 proteins only seems identical to intact viral particles on cryoelectron microscopy (Hagensee et al.,1994).

2.4 Genomic organization

The coding HPV DNA is composed of 8 ORFs all located on the same strand of DNA, and classified as *early* (E) or *late* (L) depending on their location on the genome and the stage of infection at which their associated proteins are produced. Genes from the E ORFs are expressed in non-productively infected cells and in transformed cells, and code for proteins associated with genomic replication and control. Genes from the L ORFs are expressed in productively infected cells, and code for structural proteins of the capsid (Baker and Howley,1987). In each papillomavirus genome, there is a region free of ORFs. This region is known under several names: *long control region* (LCR), *upstream regulatory region* (URR), or the *non-coding region* (NCR). Figure 1 shows graphically the genomic organization of HPVs (adapted from Scheffner et al.,1994).

The proteins encoded by the E and L ORFs will be described in the next section. The LCR region of the HPV genome contains promoters and enhancers that regulate viral transcription.

A number of regulatory proteins bind to specific sequences of the LCR, one of which is the E2 viral protein (McBride et al.,1991). Other specific cellular factors, not restricted to epithelial cells, have been proposed to intervene in the complex process controlling viral transcription of HPV (reviewed in Hoppe-Seyler and Butz,1994). The greater efficiency for immortalization seen in some HPV types (for example HPV-18 compared to HPV-16) has been attributed to differences in their respective LCR sequences (McBride et al.,1991).

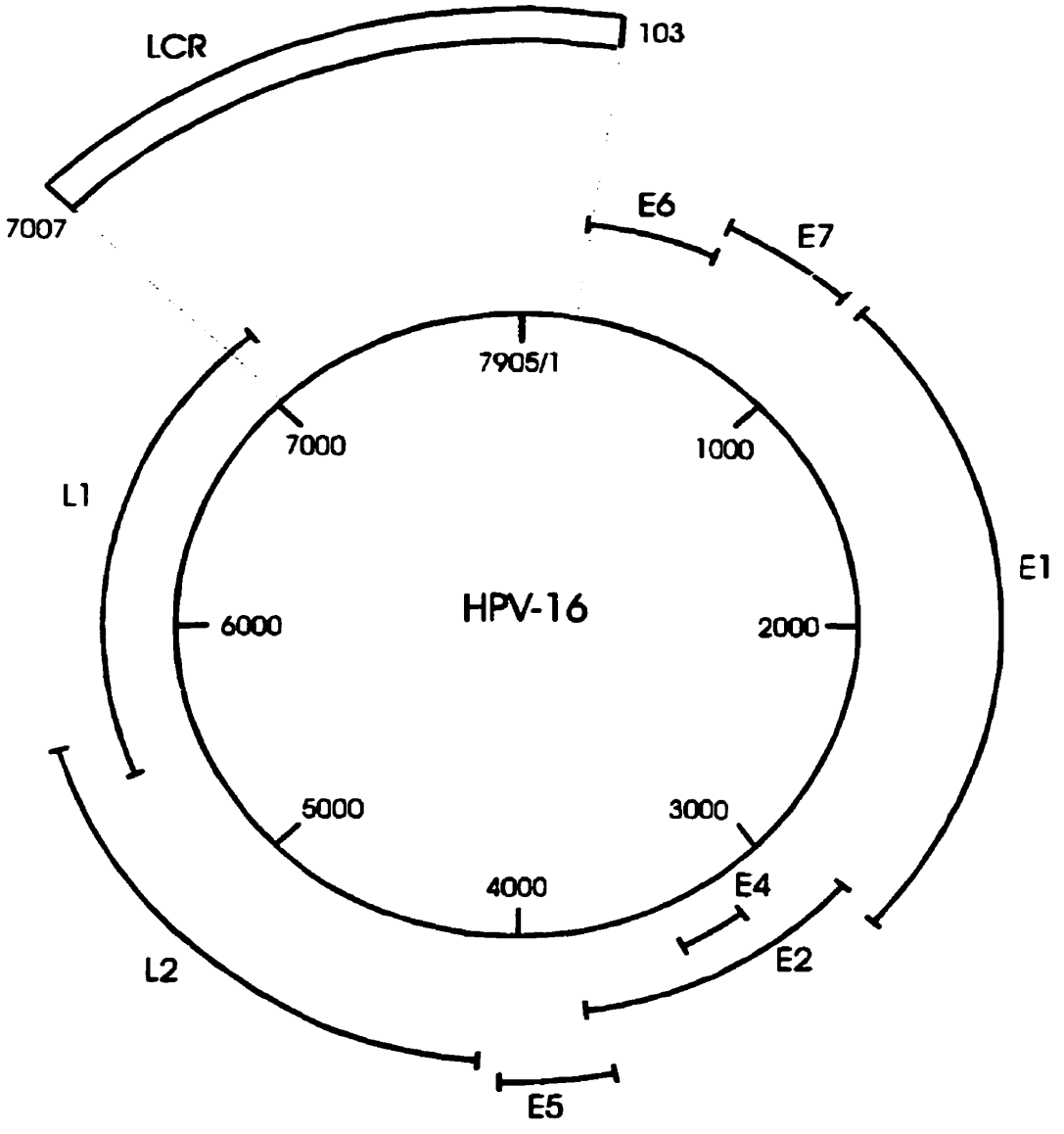


Figure 1: **Map of the HPV-16 genome**

The Early (E) open reading frames (ORFs), Late (L) ORFs, and the long control region (LCR) are identified on this map of the HPV-16 genome (adapted from Scheffner et al.,1994).

2.5 Viral replication

2.5.1 Relationship with cell differentiation

Transcription and replication of HPV are highly dependent on the level of differentiation of epithelial cells in which they are present. Only the basal cells of the squamous epithelium can be infected by HPV, since they are the dividing cells of the epithelium. Viral DNA was shown by *in situ* hybridization in basal and parabasal cells of papilloma (Schneider et al.,1987). However, even though the infection is initiated at the basal layer, the production of viral particles, genome amplification, capsid protein synthesis, and virion assembly are done in well-differentiated epithelial cells (Laimins,1993). In summary, translation of early proteins is carried out in basal and suprabasal epithelial cells, translation of late proteins and production of viral particles is undertaken only when cells infected with HPV attain a higher level of differentiation during the normal process of differentiation of a squamous epithelium (Howley,1996).

2.5.2 Attachment

The attachment of a viral particle to a host cell seems related to the L1 capsid protein. Studies of virus-like particles (VLPs) have shown an inhibition of virus attachment when VLPs made of L1 proteins were present, suggesting a competition of VLPs with the actual virions for attachment to the host cells (Roden et al.,1994). However, cell surface receptors for papillomaviruses have not yet been identified. Studies with bovine papillomavirus type 1 (BPV 1) and HPV 16 VLPs suggest a common and well conserved cell surface receptor (Roden et al.,1994). To date, there is no evidence on mechanisms of virus entry into cells, access to the nucleus, or uncoating (Howley,1996).

2.5.3 Transcription

Transcription of viral DNA normally initiates with the binding of an RNA polymerase complex to a promoter site of the genome. Transcription of papillomaviruses is complex and involves "multiple promoters, alternate and multiple splice patterns, and the differential

production of mRNA species in different cells" (Howley, 1996). One transcriptional promoter, P_{97}, has been identified in the HPV 16 genome, and is located within the LCR, upstream to the E6-E7 ORFs (Cripe et al.,1987). In HPV-18, a similar promoter, P_{105}, was found. The regulation of HPV transcription is precise, and varies between different levels of differentiation of the host epithelium (Chow et al.,1987). Viral gene expression is modulated in part by the viral genome itself that contains *cis* regulatory elements (or enhancers) and encodes for transcriptional regulatory factors. The LCR region of papillomaviruses genome includes enhancer elements that can also be activated in response to cellular factors (Cripe et al.,1987; Thierry et al.,1987). Constitutive enhancer elements are thought to be essential for initiating the expression of viral genes. They may also play a role in viral latency (Howley,1996).

2.6 HPV gene products

Table 2 summarizes the general functions of the proteins encoded by the HPV genome.

Table 2: Functions of the HPV gene products

Gene	Function of gene products
E1	Repressor of viral DNA replication
E2	Modulation of gene transcription and replication (+/-)
E4	Release of virion particles
E5	Growth stimulation
E6	Growth stimulation Association with, and degradation of p53 tumour-suppressor protein resulting in disruption of cell cycle
E7	Growth stimulation Association with Rb protein and disruption of cell cycle
L1	Major capsid protein
L2	Minor capsid protein

2.6.1 E1 protein

E1 is the largest ORF of the HPV genome and is well conserved among papillomaviruses (Chan et al.,1992). Most of the knowledge of E1 functions come from studies of BPV-1 E1 proteins. The BPV-1 E1 protein is a nuclear phosphoprotein and binds specifically to the site of origin of replication in the viral LCR (Wilson and Ludes-Meyers,1991). E1 is therefore implicated in viral DNA replication functions. The weak binding affinity of the E1 protein to the viral LCR is increased by the binding of the E2 protein (Ustav et al.,1991; Mohr et al.,1990). Studies of induced mutations in the HPV-16 E1 gene have resulted in increased immortalization capacity of the viral genome, suggesting that E1 may be able to repress the P_{97} promoter, by a yet unspecified mechanism (Romanczuk and Howley,1992). The E1 protein would therefore down-regulate replicative activity of the HPV genome. Consistent with this finding is the fact that integrated HPV DNA in cervical cancer normally shows intact E6/E7 ORFs while E1 and/or E2 are disrupted, with their respective proteins not produced (Scheffner et al.,1994).

2.6.2 E2 proteins

Proteins coded by the E2 gene are involved in the regulation of viral transcription and replication. The E2 gene products can bind to specific sites of the LCR of the viral genome and act as transcriptional activators and repressors or as modulators of replication (McBride et al.,1991). The HPV-16 and HPV-18 respective major promoter, P_{97} and P_{105}, are negatively regulated by E2 (Thierry et al.,1987; Bernard et al.,1989; Romanczuk et al.,1990). The down-regulation effected by E2 is mediated through binding of E2 to two E2-binding sites immediately adjacent to the promoter sites P_{97} and P_{105} (Romanczuk et al.,1990). Binding of E2 to these sites is thought to interfere with the initiation of viral transcription.

In most cervical tumours, the HPV genome is integrated into host cell DNA with disruption of E1 and/or E2 ORFs. Down-regulation by E2 would be ceased by a disruption in the E2 ORF as a result an over-expression of E6 and E7 proteins. Such proteins would bind to p53 and Rb gene products, resulting in an alteration of the normal cell cycle and leading to

uncontrolled proliferation (Werness et al.,1990; Scheffner et al.,1990; Dyson et al.,1989; Münger et al.,1989). Since the predominant effect of the E2 protein is to down-regulate the expression of HPV E6 and E7 genes, loss of E2 protein expression would be expected to enhance the immortalization capacity of the virus, which was confirmed by studies with mutated HPV E2 (Romanczuk and Howley,1992).

2.6.3 E4 protein

The E4 gene is located in the early portion of the genome, but is expressed late in the viral life cycle, only in the most differentiated cells (Doorbar and Gallimore,1987). mRNAs translated from viral E4 are among the most abundant transcripts in biopsies of cervical intraepithelial neoplasia and condyloma (Crum et al.,1988). The major product of the E4 gene is a 17Kd polypeptide whose cleavage or dimerisation leads to other E4 coded proteins. Proteins encoded by E4 are primarily found in the cytoplasm of the host cell (Crum et al.,1990). More specifically, if expressed in human keratinocytes, the HPV E4 protein is localized in cytoplasmic inclusion granules, and causes the collapse of the epithelial cell intermediate filament network that leads to the release of viral particles (Doorbar et al.,1991). E4, therefore, seems to play a role in the cytopathic effects produced by HPVs and is closely linked to the productive stages of the virus life cycle.

2.6.4 E5 protein

The E5 gene product is a hydrophobic protein of 10 kDa found within the infected cell's Golgi membrane. Studies of the function of E5 have mostly been done in BPV. In rodent cell cultures, E5 has been found to be one of the major oncogenic transforming proteins (Schiller et al.,1984). The transforming activity attributed to BPV E5 is mostly mediated by inhibition of the normal degradation of cell surface receptors for cellular growth factors (Velu et al.,1987). In the case of HPV-16, the E5 protein has been shown to stimulate growth in cells expressing high levels of epidermal growth factor (EGF) receptors (Pim et al.,1992). It had been hypothesized that the effect of the HPV E5 protein may result from an abnormal amplification of growth stimulatory signals from growth factor receptors (Leechanachai et

al.,1992). The augmented growth potential due to the E5 gene product has been associated with its binding to the 16Kd component of the vacuolar ATPase and a resulting inhibition of acidification of endosomes in human keratinocytes (Conrad et al.,1993; Straight et al.,1995). These findings suggest a model in which the E5 protein binds to the 16Kd component, impairing its activity and therefore resulting in a less efficient degradation of EGF receptors. Other activities of the E5 protein have been suggested and are currently being investigated. The multifunctional role of the E5 gene product would include enhancement of signal transduction, reduction in endosome activity, upregulation of viral gene expression, impairment of antigen presentation and of cell communications (Banks and Matlashewski,1996).

2.6.5 E6 and E7, the most important transforming proteins

The critical proteins for development of malignant cell characteristics are encoded by the E6 and E7 ORFs. When the viral genome is integrated to the host DNA in malignant lesions, the expression of E6 and E7 genes is maintained and even increased. Biological studies with transgenic mice showed that the expression of E6 or E7 alone resulted in development of condyloma, intraepithelial neoplasia and invasive squamous cell cancer on the skin of the mice (Lambert et al.,1993). Some important discrepancies found between the E6 and E7 proteins of high- and low-risk HPV types might explain the drastically different behaviour of these types in terms of induction of malignancy. These discrepancies are: 1) tissues infected with high-risk HPV types show a higher expression of E7 than tissues infected with low-risk types; 2) E6 and E7 of high-risk HPV types show a much stronger affinity for the tumour-suppressor proteins and inactivate them more efficiently leading to a substantial disruption of the normal cell cycle (Smotkin et al.,1989; Sedman et al.,1991).

The tumour-suppressor proteins to which HPV E6 and E7 gene products show affinity are the p53 and Rb proteins. The p53 protein has been named the "guardian of the genome", in recognition of its ability to preserve genomic integrity (Lane,1992). The p53 gene product has been shown to be an important negative regulator of cell growth and a tumour-suppressor

protein. In response to genomic alterations that would be passed to the next cell generation, the p53 protein can arrest the cell cycle in the G1 or G2 for DNA repair mechanisms to take place, or otherwise induce apoptosis, also known as programmed cell death (Shaw et al.,1992). Another important tumour-suppressor protein, the retinoblastoma (Rb) gene product, acts as a negative regulator of cell growth via an inhibition of the effect of positive growth regulators (Dyson et al.,1989).

E6 is an approximately 150-amino acid protein, not capable of inducing malignant transformation alone, but has been shown to immortalize keratinocytes effectively in combination with E7 (Hawley-Nelson et al.,1989). It was experimentally shown that E6 had the affinity to bind the p53 tumour-suppressor protein (Werness et al.,1990) which lead to its degradation via an ATP-dependent mechanism involving the ubiquitin-dependent protease system (Scheffner et al.,1990). This degradation of the p53 protein is greater when associated with HPV-16 E6 gene product than with HPV-18's. It does not occur with HPV-6 or HPV-11 E6 proteins (Scheffner et al.,1990). The binding of the E6 gene product to the p53 protein therefore disrupts the normal response of cells to DNA damage and may increase genetic instability in HPV-infected cells (Kessis et al.,1993).

The protein encoded by the E7 ORF is a 98-amino acid protein found in the cytoplasm of the host cells (Smotkin and Wettstein,1987). The expression of E7 alone is sufficient for malignant transformation, but cell growth is more efficient when E6 is also expressed (Halbert et al.,1991). The HPV E7 protein was shown experimentally to bind the Rb protein (Dyson et al.,1989), another important negative regulator of cell growth such as p53, and to inactivate it (Münger et al.,1989), therefore preventing the Rb protein to inhibit uncontrolled growth. The E7 gene product may also act more directly by interacting with proteins involved in the normal cell cycle control such as the protein kinase $p33^{CDK2}$ and cyclin A (Tommasino et al.,1993).

2.6.6 L1 and L2, the capsid proteins

Two structural proteins compose the viral capsid: L1 (57 kDa) the major capsid protein and L2 (77kDa), the minor capsid protein (Doorbar and Gallimore, 1987). L1 is the most abundant one, accounting for approximately 80% of the protein content of the viral capsid (Favre et al., 1975). L1 is a highly conserved protein among the different papillomaviruses types, and contains cysteine residues, also highly conserved, at six positions. These cysteine residues form disulfide bonds only between L1 proteins, which does not seem to be linked to the minor capsid protein in this way (Doorbar and Gallimore, 1987). Expression of L1 and L2 genes of HPV is strongly associated with epithelial cell differentiation, with the highest levels of L1 and L2 expression seen in well-differentiated cells.

2.7 Integration of HPV DNA in host genome

In carcinoma *in situ* and invasive cancer of the cervix, the HPV viral genome is found integrated, and there is no production of viral particles (Laimins, 1993). The integration frequently disrupts the E1 and/or the E2 ORFs, but leaves intact the LCR, and the E6/E7 ORFs. The disruption of E2 and resulting suppression of its down-regulating effect on transcription of E6 and E7 genes results in increased transcription of E6 and E7 (Sang and Barbosa, 1992), followed by their respective effect on cell cycle control by their binding of p53 and Rb proteins. Recent results tend to confirm that integration of the HPV DNA into the host's genome is a crucial event in the progression from high-grade pre-neoplastic lesion to neoplastic tumour (Daniel et al., 1997). This recent study also confirms the E2 ORF disruption theory. Polymerase chain reaction (PCR) analysis and RNA *in situ* hybridization have resulted in amplification of a gene segment including the E2-E5-L2 region of the HPV genome in 15 out of 16 CIN 3 lesions, and in no amplification of this segment in all 19 tumours analysed. The same analysis with a fragment comprising the E6-E7 region of the genome was positive in all 16 lesions and 19 tumours (Daniel et al., 1997). The analysis of stable cervical cancer cell lines has demonstrated that integration can occur at various sites in the host genome, but frequently occurs close to proto-oncogenes resulting in increased expression (Franco, 1993). Study of the site of integration in cervical biopsies from squamous cell carcinoma and CIN

3 lesions has suggested that integration could take place at different sites and appeared to be a random event (Matulic and Soric,1994).

3 Detection of HPV infections

3.1 Morphology

The presence of HPV can be indicated by its cytopathological effect on the cervical epithelium. However, koilocytosis, which is a pathognomonic feature of HPV infection is seen almost exclusively in productive viral infections and not in cervical cancer precursors, which precludes its use in determining potentially malignant lesions. Such morphological detection does not allow identification of the genotype of HPV present which requires molecular biology techniques. Since the vast majority of HPV infections do not result in abnormal cellular changes, detection based on morphological abnormalities does not correlate well with the presence of HPV DNA in the cervical epithelium. For example, in a study among university students where detection of HPV DNA was performed by molecular hybridization techniques, 30% of the women with normal cytologies harbored cervical HPV DNA (Bauer et al.,1991). These women would have been false negatives if the presence of cervical HPV had been determined based on morphology.

3.2 Detection of HPV DNA

The detection of HPV DNA in cervical cell specimens by molecular hybridization techniques is the method of choice, since it allows the identification of latent, subclinical, and clinical infections.

Over the years, many techniques with a wide range of specificity, sensitivity, and reproducibility have been used to detect HPV. Some of the early techniques used have resulted in misclassification of HPV status, biasing the association between HPV infection and cervical neoplasia towards the null (Franco,1991).

Southern blot was one of the first methods used in small studies but is very labour intensive, and has since been replaced with other techniques with the advent of large epidemiological studies. It requires purification of DNA, digestion with restriction enzymes, electrophoresis of the fragments produced, their denaturation and transfer to a filter. Hybridization is then performed with radioactive probes in different stringency conditions. The size of fragments and stringency of hybridization are used to identify the HPV types (IARC Working Group, 1995). This technique requires a large amount of cellular material, is labour intensive, and demands expertise in order to obtain reproducible results (Schiffman, 1992; Johnson, 1995).

Filter *in situ* hybridization (FISH) was the first HPV test developed for large scale epidemiological studies. It was used in various studies (Reeves et al., 1989; Villa and Franco, 1989; Kjaer et al., 1990), but resulted mainly in misclassification and bias of the estimates for HPV status and cervical neoplasia (Franco, 1991). For this technique, cervical cells are placed on a filter, lysed, and DNA is hybridized with radioactive probes. An autoradiograph allows identification of the HPV types present. When compared to Southern blot, a knowingly imperfect gold standard due to problems with reproducibility, virtually no correlation was found between the two assays (Schiffman, 1992).

One of the first commercialized assays for clinical HPV testing was the ViraPap™/ViraType™ kit (Digene Laboratories, Silver Spring, MD, USA). It has been approved by the American Food and Drug Administration (USFDA). Technically, this method consists of placing the cellular DNA from cervical specimens onto a filter. The DNA is then hybridized with 7 RNA probes for the most common HPV types (6, 11, 16, 18, 31, 33, 35). An autoradiograph allows revelation of the hybrids in dot blot format (IARC Working Group, 1995). Validation studies of this method by comparison with Southern blot have shown comparable sensitivity and specificity (Kiviat et al., 1990; Moscicki et al., 1993). The major drawback is the restricted number of types that can be identified. Digene Laboratories has now replaced ViraPap™/ViraType™ by an assay called Hybrid Capture™. Advantages

of this test are the increased number of RNA HPV probes (14), and the non-radioactive method of detection. A hybrid-specific antibody recognizes the RNA-DNA hybrids, and detection is based on quantification of a chemiluminescent reaction (IARC Working Group, 1995).

Polymerase chain reaction-based methods (PCR) are often used for HPV testing in epidemiological studies. These methods have the advantage of requiring very little cellular material for detection, and are highly sensitive and specific. In order to amplify at the same time HPV from various types, primers that targeted highly-conserved regions of the genome were designed. The resulting amplification products are then either typed with type-specific oligonucleotide probes or analysed by restriction fragment length polymorphism (RFLP). The high sensitivity of techniques based on PCR allows detection of HPV DNA in samples that contain as little as 10-100 HPV genomes (IARC Working Group, 1995). The two most widely known PCR protocols for HPV amplification and detection are the MY09/MY11 and the GP5+/GP6+, initially developed by Manos et al. (Manos et al., 1989) and Van den Brule et al. (Van den Brule et al., 1990) respectively . The MY09/MY11 protocol amplifies a highly-conserved 450-base pair (bp) fragment of the L1 gene. This protocol has been modified to amplify and detect HPV type 51 (Hildesheim et al., 1994). The GP5+/GP6+ protocol targets a 140-bp region of the L1 gene. It has been modified to improve its performance (Husman et al., 1995). Generic probes and type-specific probes for more than 25 genital types are commonly used. The main feature of the PCR-based technique, its high sensitivity, is also one of its major drawbacks. Great care and strict laboratory procedures have to be observed to prevent contamination.

4 Pathogenesis of human papillomaviruses

4.1 Transmission

The sexually-transmitted profile of HPV was first recognized for non-oncogenic types, mostly types 6 and 11, which result in condyloma acuminata or genital warts. In 1954, Barrett et al. reported sexual transmission of genital warts among wives of men who had previously experienced genital warts (IARC Working Group,1995). Studies of young women with no previous sexual experience have shown that HPV DNA can be absent (Fairley et al.,1992; Rylander et al.,1994), but can also be present in some cases (Pao et al.,1993). Studies of sexually active couples have shown same type HPV infections among partners to be significantly higher than expected (Baken et al.,1995). However, same types and variants of HPV are not necessarily detected in genital samples from both sexual partners, suggesting sexual transmission, but with low infectivity (Ho et al.,1993). Using PCR as a detection method, the association between markers of sexual activity and HPV prevalence has varied in different studies from strong (Ley et al.,1991; Bauer et al.,1993; Wheeler et al.,1993), moderate (Rohan et al.,1991; Hildesheim et al.,1993), to non-existent (Kjaer et al.,1993). Results from a Northeastern Brazil study among cytologically normal women showed evidence suggesting that transmission of high-risk HPV types may depend more strongly on sexual activity than for low-risk types (Franco et al.,1995). Sexual contact is still considered the main mode of transmission, but other less important nonsexual routes should also be considered and investigated more thoroughly.

Possible modes of transmission in children include sexual abuse, perinatal contamination, autoinoculation, fomite transmission and nonsexual contact (Gutman et al.,1993). Sexual abuse remains the most plausible mode of transmission for anal, genital, and oral HPV infections in children, as many children found to have HPV-related anogenital diseases also had a history of sexual abuse (Gutman et al.,1993; Gutman et al.,1994). Vertical transmission of HPV is considered as highly plausible. A recent study showed presence of oral HPV DNA in one-third of children born from mothers with a history of genital HPV infections, with half

of the children harbouring the same HPV type as their mother (Puranen et al.,1996). The other possible modes of transmission to children: autoinoculation most likely from hand warts, fomite transmission, and nonsexual intimate contact seem to be of lesser importance, and have not been well documented (Puranen et al.,1996).

4.2 Type of infection and morphological abnormalities

Presence of cervical HPV DNA can lead to many different clinical and subclinical outcomes. The infection itself, or presence of the virus, can be cleared apparently by the host's immune system. If not cleared, the virus can persist causing little or no cytological abnormalities. If cellular abnormalities are caused, they can be transient and eventually regress. In certain cases, cytological abnormalities will persist, even progress to more severe clinical signs that can also regress, progress or remain as such. The non-linear continuum of cellular changes mediated by HPV results in different levels of abnormalities. Infection with HPV can therefore be clinical, subclinical, or latent. Clinical infections refer to those morphological changes causing symptoms that are visible to the unaided eye. Subclinical infections do not easily result in visible symptoms and can only be diagnosed with colposcopy, cytology or histology. Latent infections are those which result from presence of the virus in cervical epithelial cells, without morphological abnormalities to the epithelium (Schneider and Koutsky,1992).

Koilocytosis is a distinctive characteristic of cells infected with HPV. Koilocytosis consists of the combination of nuclear atypia and perinuclear halo. Nuclear atypia consists of enlargement, hyperchromasia, and irregularity of the nuclear membrane (Wright et al.,1994). The perinuclear halo is thought to be caused by the expression of the viral E4 protein which results in the collapse of the host cell's cytokeratine matrix, probably facilitating the exit of virions (Doorbar et al.,1991). LSILs are morphologically characterized by these features of productive HPV infection. Nuclear abnormalities must be present for the cytomorphological effects to be attributed to HPV (Wright et al.,1994). Less than 25% of the lower part of the epithelium shows nuclear atypia, few mitoses can be seen and none of them are abnormal (Palefsky and Holly,1995).

In HSIL, the more superficial layers of the epithelium contain cells with abnormal nuclei and cells with a high nuclear:cytoplasmic ratio can be found in the lower portion or throughout the epithelium. Dividing cells can be found in all layers (Palefsky and Holly,1995).

Other benign clinical conditions have been related to genital HPVs, such as genital warts, laryngeal papillomas and papillomas at other mucosal sites (Shah and Howley,1996). Genital warts are caused by HPV 6 and 11, two types classified as low-oncogenic HPV types.

4.3 Immune response to papillomavirus infections

The humoral response to HPV results in production of IgM, IgG, and IgA, sometimes simultaneously, and in some cases with a delayed appearance of IgG or IgA. IgM and IgA responses seem to decrease after clearance of infection, whereas IgG responses persist (Wikström et al.,1995). Serological studies have shown that many subjects who harboured HPV infections did not have measurable levels of antibodies against HPV. The development of an antibody response to HPV is thought to be slow and short-lasting likely caused by the intraepithelial location of infection and the resulting low exposure to the immune system. Seropositivity is associated with prolonged exposure to capsid antigens, suggesting a less likely antibody response in transient infections (Galloway,1996). Serology is a measure of overall exposure to HPV antigens, and is therefore not restricted to genital HPV types that are of interest in the study of cervical cancer.

The cell-mediated response to HPV infection appears to be important in clearing infections and limiting the size of lesions. Immunosuppressed renal-transplant and HIV infected patients who have impaired cell-mediated immunity present HPV-related lesions that grow larger, show less regression, and persist longer. However, since papillomaviruses infect only epithelial cells and do not lyse them, little or no local inflammation is produced resulting in a weak cell-mediated immune response (Frazer and Tindle,1996). Abundant CD8+ infiltrate in regressive genital warts support the role for cell-mediated immunity in containment of lesions (Coleman et al.,1994).

5 Epidemiology of HPV infection

HPV infection has been studied widely, with various methodologic approaches, in a multitude of geographical areas, and in populations with varying characteristics. Most importantly, until recent years a multitude of methods were used for detection of HPV DNA in cervical cell samples and showed limited comparability (Schiffman, 1992). Nevertheless, these results have allowed epidemiologists to compile crucial information helpful to the understanding of the natural history of genital HPV infection and cervical neoplasia. Recent studies have used PCR-based methods for detection of HPV DNA due to its higher sensitivity and specificity.

5.1 Occurrence of infections

The prevalence of HPV infection in cytologically normal women, as derived from the most recent molecular epidemiology studies, is generally between 15% and 40% when detection is done with PCR-based methods. Table 3 provides a summary of selected studies that measured the prevalence of HPV infection among women without cytological abnormalities (adapted from IARC Working Group, 1995). These studies all used modern PCR-based protocols and are less subject to misclassification than earlier ones. Listed studies had reasonably large sample sizes. They are presented by decreasing prevalence, but many factors make them difficult to compare: use of different PCR-based protocols, different population characteristics, different geographical areas, and different cytological grading systems.

These studies have provided very important insights into the natural history of genital HPV infection. Prevalence of infection is related to age, with a higher prevalence among women in their early twenties which decreases with age (Ley et al., 1991; Morrison et al., 1991; van den Brule et al., 1991). Young women who report being virgins have much lower prevalence rates of cervical HPV: 3.6% among American university students 18-20 yrs of age (Critchlow and Koutsky, 1995), 1.5% from a Swedish adolescent clinic (Rylander et al., 1994), and 0% from an Australian clinic with a population aged 13-41 who reported no previous penetration (Fairley et al., 1992). In terms of distribution of HPV types, HPV 16 is the most prevalent type in most studies among cytologically normal women (Schiffman, 1994).

Table 3: Prevalence of HPV DNA detection in cervical samples from women with normal Pap smears - Selected studies

Reference	Country, Study population	Sample size	HPV types tested	% HPV positive
(Wheeler et al.,1993)	USA, University Health Service	357	6, 11, 16, 18, 31, 33, 35, 39, 45, 51, 52, 54, 59, PAP88, PAP238A, W13B	44.3
(Kjaer et al.,1993)	Greenland, Random sample	129	11, 16, 18, 33	43.4
(Kjaer et al.,1993)	Denmark, Random sample	126	11, 16, 18, 33	38.9
(Critchlow and Koutsky,1995)	USA, University	183	6, 11, 16, 18, 31, 33, 35, 45	35.0
(Hildesheim et al.,1993)	USA, Screening	404	6, 11, 16, 18, 26, 31, 33, 35, 39, 45, 51-59, 68, W13B, PAP88, PAP155, PAP238A, PAP291	33.7
(Bauer et al.,1991)	USA, University Health Service	442	6, 11, 16, 18, 31, 33, 35, 39, 45, 51, 52, W13B, PAP88, PAP155, PAP251, PAP238B	32.5
(Richardson,1996)	Canada, University clinic	449	6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 45, 51-59, 66, 68, PAP155, W13B, PAP238A, PAP291	21.8
(Franco et al.,1995)	Brazil, Screening	525	6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 45, 51-59, 66, 68, PAP155, W13B, PAP238A, PAP291	18.3
(Rohan et al.,1991)	Canada, University Health Service	105	6, 11, 16, 18, 33	18.1
(Schiffman et al.,1993)	USA, Health Maintenance Organization clinics	453	6, 11 16, 18, 26, 31, 33, 35, 39, 40, 42, 45, 51, 52, 53, 54, 55, 57, 59, PAP38, PAP155, PAP238A, PAP251, PAP291, W13B	17.7
(Melkert et al.,1993)	Netherlands, Screening 15-34 yrs	156	6, 11, 16, 18, 31, 33	14.1
(van Doornum et al.,1992)	Netherlands, STD clinic	108	6, 11, 16, 18, 33	11.9
(Bosch et al.,1993)	Colombia, Random sample	181	6, 11, 16, 18, 31, 33, 35	10.5
(Bosch et al.,1993)	Spain, Random sample	193	6, 11, 16, 18, 31, 33, 35	4.7
(Melkert et al.,1993)	Netherlands, Screening 35-55 yrs	1555	6, 11, 16, 18, 31, 33	4.2

Most of the studies were based on cross-sectional designs. Very few studies have been designed to estimate the incidence of genital infection since this requires a longitudinal follow-up and multiple testing of HPV DNA over time. A prospective study of heterosexual men and women with multiple sexual partners having HIV as a primary endpoint of interest, also measured HPV status during follow-up from May 1988 to January 1991 (van Doornum et al.,1994). This study is, to our knowledge, the only study that provided incidence-density rates of HPV infection using person-time information as a denominator. However, the incidence rates calculated include HPV detection from genital, oral, and anorectal regions. Such incidence for women was found to be 47.1 infections per 100 person-years, based on 99 out of 110 women who developed HPV infection during follow-up, 59 (60%) of which were genital infections (van Doornum et al.,1994).

5.2 Determinants of prevalent infections

HPV infection is seen today as the intermediate step in the causal pathway of cervical neoplasia, between sexual activity and precursor lesions. Its determinants are therefore likely to show a high similarity with those of cervical cancer. This similarity even goes beyond the association with markers of sexual activity.

With the advent of PCR-based methods, the sexually-transmitted profile of HPV infection has been revealed (Ley et al.,1991). As mentioned in the discussion of risk factors for cervical cancer, the two main markers of sexual activity or behavior that have been studied are age at first intercourse, and more importantly number of sexual partners. Even with misclassification of HPV status being less of a problem due to the use of sensitive and specific detection methods, the association between markers of sexual activity and HPV infection has been controversial (Ley et al.,1991; Bauer et al.,1993; Wheeler et al.,1993; Rohan et al.,1991; Hildesheim et al.,1993; Kjaer et al.,1993). A recent study in Northeastern Brazil has shown a different association of HPV infection and markers of sexual behaviour according to oncogenic potential. Detection of high-risk HPV types was strongly associated with sexual activity, whereas a weak correlation of low-risk infection and sexual activity was seen only

in women below 40 years of age (Franco et al., 1995). Similar findings have been obtained in a study among university students in Montreal, where prevalence of high- but not low-risk HPV types was associated with markers of sexual activity (Richardson, 1996).

Age is the second most important determinant of HPV infection, with a high prevalence seen for women in their early twenties, and a sharp decrease afterwards (Ley et al., 1991; Morrison et al., 1991; van den Brule et al., 1991). The association of age with infection is maintained after adjustment for other risk factors, indicating independent effect (Ley et al., 1991; Bauer et al., 1993; Wheeler et al., 1993). The hypothetical explanation of this effect is by acquired immunity or hormonal changes due to increasing age (Schneider, 1996).

Immunosuppression is a determinant of prevalence of infections, with a marked higher prevalence of HPV genital infections among HIV positive subjects compared to HPV negative individuals. Prevalence of genital HPV infection was found to be 2 times and 6 times higher in HIV positive individuals compared to HIV negative individuals respectively by Williams et al. (1994) and Wright et al. (1994) (reviewed in Schneider, 1996).

Other suggested independent risk factors are smoking, oral contraceptive use, parity, and diet (Bauer et al., 1993; Rohan et al., 1991; Wheeler et al., 1993). Smoking may act either via a direct carcinogenic or promoting effect of nicotine metabolites on the cervical epithelium (Schiffman et al., 1987) or in an immunosuppressive manner, reducing the hosts ability to clear the infection (Barton et al., 1988). A potential explanation for the role of oral contraceptive use and parity in acquisition of genital HPV infection would be through the negative hormonal influence on the immune response to the virus. HPV infection tends to be detected with an increased prevalence among pregnant women according to Schneider et al. (1987) (reviewed in Franco, 1993). Multiple pregnancies could hypothetically increase the likelihood of HPV infection in providing many periods with decreased immunity and physical trauma of the cervical epithelium (Franco, 1993). A protective effect of vitamins A and C and

β-carotene is suspected, given their protective role in cervical cancer (Herrero et al.,1991). However, their effect on HPV infection per se remains to be studied.

To date, no results have been published on risk factors for incident infections.

AIMS OF THE STUDY

The present study is designed to focus on incident HPV infections of the cervix among a population from a high-risk area for cervical cancer. This study on the incidence of cervical HPV infections is one the first to use cervical HPV status longitudinally, at regular time intervals. The aims of the project are as follows:

- To calculate the incidence density rates for overall, low-risk, and high-risk cervical HPV infections.

- To identify the determinants of incident infections for low-risk, and high-risk HPV types.

1 Study design

1.1 Overview

This project uses data collected from the prospective Ludwig-McGill Cohort Study on the molecular epidemiology and natural history of cervical HPV infection and neoplasia. The main purpose of this cohort is to study persistent HPV infection, particularly with high-risk types, and its subsequent role in the development of cervical lesions. The study was initiated in November 1993 and the accrual phase ended in March 1997, with over 2500 women recruited and to be followed for up to 5 years. Visits are scheduled every 4 months during the first year and once yearly thereafter, for a total of 8 visits.

1.2 Setting

The study is being conducted in São Paulo, Brazil, at the "Hospital Municipal Maternidade Escola Dr. Mario de Moraes Altenfelder Silva", an institution that is also locally known as "Maternidade Escola Vila Nova Cachoeirinha" (MEVNC). This hospital is part of a network of primary, secondary and tertiary health care institutions coordinated by the municipal health department.

1.3 Subject recruitment

Women were selected from the outpatient lists of the gynaecology, family medicine, and family planning clinics of the MEVNC, based on a systematic sampling selection process. Selected women are approached by research nurses who inform them about the general purpose of the study, and verify whether they meet the eligibility criteria detailed in the next section. Eligible women are then given further explanations about the study, frequency of return visits, required physical examinations, specimen collection, and interviews. Women are also informed that an incentive to compliance will be given at each visit, consisting of a meal ticket whose value ranges from US$5 at first visit to US$20 for first annual and subsequent

follow-up returns. The monetary incentive for compliance is given in order to ensure the highest return rate possible in this long-term study that requires substantial involvement on the part of its participants. Women who meet the eligibility criteria and accept the requirements are then asked to sign an informed consent form. The study was approved by the ethical review committees of McGill University, University of Toronto, Ludwig Institute for Cancer Research, and MEVNC.

1.4 Inclusion and exclusion criteria

The following inclusion and exclusion criteria were used to determine women's eligibility for the Ludwig-McGill cohort study:

i. Must be between 18 and 60 years of age inclusively. This range covers the important time-period for a natural history study on HPV infection and neoplasia of the cervix. Previous data in a similar population has shown that prevalence of infections is at its highest in women in their early twenties, incidence of carcinoma *in situ* peaks in women in their mid-thirties, and finally the highest incidence of invasive cervical cancer occurs in women 60 years of age (Mirra and Franco, 1985);

ii. Must be a permanent resident of the city of São Paulo, Brazil, in order to avoid loss of follow-up due to moving after a temporary stay in the city;

iii. Must not be pregnant at the time of recruitment, nor planning a pregnancy for the following 12 months. Pregnancy would eventually interfere with cervical cell sample collection and cervicography. This exclusion criterion aims at minimizing bias in missing information, and assuring that a large enough proportion of women enrolled will have all specimens and information collected at every visit;

iv. Must have an intact uterus and not being a referral for hysterectomy at recruitment, thereby restricting the study population to women at risk for cervical cancer;

2 Data collection

2.1 Interviews

All interviews were based on standard questionnaires and conducted by two trained research nurses unaware of the HPV status of the women. The interviews took place in private, sound-proof examination rooms. The baseline interview was comprised of questions that referred to lifetime or pre-defined time periods, whereas the return interviews probed for information specifically referring to the period of time passed since the last visit. For this particular analysis, only information from the baseline interview was used. The original Portuguese version of the questionnaire used by the research nurses can be consulted in Appendix 1. This questionnaire contains 107 questions that focus on the different areas of importance regarding cervical HPV infections and neoplasia: sociodemographic characteristics, sexual practices, reproductive health, contraception, and lifestyle factors such as smoking, alcohol drinking and hygiene. Since the questionnaire is only used in Portuguese, there exists no official French or English versions. However, Appendix 2 is a summary table of the baseline questionnaire indicating the information collected with each series of questions. The test-retest reliability of interview-derived information on sexual activity of women has been studied previously. Two studies have shown high test-retest reliability for specific aspects of sexual behaviour such as the age at first sexual intercourse, number of sexual partners, and frequency of sexual intercourse (Rohan et al., 1994; Kunin and Ames, 1981). However, these two studies have been performed on women younger and with a higher level of education than women in the present study. As more data accumulate, the information collected in the Ludwig-McGill cohort study will be used to study this aspect.

2.2 Cervical cell specimens and HPV DNA testing

At each visit, cervical cell specimens were collected with a cytobrush (Medscand Inc.) from the ectocervical and endocervical regions. The cytobrush containing the exfoliated cells was immersed in a tube containing Tris-EDTA buffer at pH 7.4 and kept for a maximum of 5 days at the clinic, at 4°C. The tubes were shipped to the laboratory at the Virology Unit of the

Ludwig Institute for Cancer Research, in São Paulo. Tubes were then vortexed to release the cells from the cytobrush, and the resulting cell suspensions were frozen at -70°C before being tested. Initially, all specimens were systematically analysed in the order they were collected. Towards the end of the data collection for the present project, priority was given to the testing of multiple specimens from the same women. To this end, the women first enrolled in the study were identified and their second, third and fourth specimens were analysed before those of participants coming for the first time. This procedure allowed to obtain a complete year of follow-up for as many women as possible (HPV status for 4 visits).

Specimens were thawed, cells were digested, and DNA was extracted and then purified following standard techniques. Specimens were tested for the presence of HPV DNA using a previously described polymerase chain reaction (PCR) protocol. Briefly, a highly conserved 450-bp segment of the L1 viral gene flanked by primers MY09/11 (Bauer et al., 1991) was amplified. An additional primer, HMB01 (Hildesheim et al., 1994) was included in the primer mixture to allow amplification of HPV 51 specifically. A dot blot hybridization was then performed with the amplified products using oligonucleotide probes to identify the HPV types present. HPV positivity was first confirmed using a generic probe containing L1 fragments from HPV 16, 18, 51, and 66. Type-specific oligonucleotide probes were then used to identify the HPV types for which sequences of the type-specific probes within the MY09/11 fragment have been published. Twenty-four of these HPV types have taxonomic names available and can be listed as follows: types 6/11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 45, 51, 52, 53, 54, 55, 56, 57, 58, 59, 66, 68, and 73. Type-specific probes were also used for four additional types which do not yet have taxonomic classification: PAP155, PAP238A, PAP291, W13B (Hildesheim et al., 1994). For specimens that gave positive results with the generic probe, but not with any of the type-specific probes, further analyses were done. Amplified products of the L1 fragment were subjected to restriction fragment length polymorphism (RFLP), that allows to identify additional HPV types based on patterns observed when cleaving the amplified DNA with a set of restriction enzymes (Bernard et al., 1994). This additional testing allowed identification of 10 official types and 4 novel ones not yet classified taxonomically:

HPV types 32, 34, 44, 61, 62, 64, 67, 69, 70, 72, and IS39 (Peyton et al.,1994), CP6108, CP8304, and CP8061 (Peyton and Wheeler,1994). Specimens that could not be identified as known types using dot blot, type-specific probes or RFLP were considered as an unknown type and classified into the low-risk group. The study population was therefore tested for a spectrum of more than 40 genital HPV types. HPV types were classified into 2 groups based on oncogenicity for the purpose of this study: low-risk and high-risk types (Bauer et al.,1993). Low-risk HPV include types 6, 11, 26, 32, 34, 40, 42, 44, 53, 54, 55, 57, 59, 61, 62, 64, 66, 67, 68, 69, 70, 72, 73, PAP155, PAP291, W13B, IS39, CP6108, CP8304, CP8061, and unknown types. The following are classified as high-risk HPV types: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, and 58. For the present study, the term "HPV infection" refers to the detection of HPV DNA based on the previously described method.

Integrity of the specimen and presence of human DNA was verified for each sample with the amplification of a 268-base pair region of the human β-globin gene using primers GH20 and PC04 (Bauer et al.,1991). Results were visualized under ultraviolet light following gel electrophoresis and staining with ethidium bromide.

Test-retest reliability of HPV DNA detection with the L1 consensus primer polymerase chain reaction technique has been studied and shown to be high. One study has demonstrated test-retest agreement for 98% of the specimens containing adequate DNA (Schiffman and Schatzkin, 1994). PCR-based protocols have the highest sensitivity and specificity among HPV DNA detection methods (IARC Working Group, 1995).

Standard laboratory measures were taken in order to prevent any contamination of reagents or specimens including: processing of samples and laboratory testing in separate rooms, use of disposable gloves, autoclaved solutions, aerosol-free tips, and highly diluted positive controls. Negative controls, containing no HPV DNA, were processed in every set of samples in order to monitor closely the possibility of contamination.

v. Must not have used vaginal medication during the 2 days before the recruitment visit. Recent use of vaginal medication hampers the ability to collect good quality cervical cell specimens;

vi. Must not have been treated for cervical disease by electrocoagulation, cryotherapy, or conization, the methods used at MEVNC. This criterion eliminates women already diagnosed with one of the events of interest (pre-neoplastic or neoplastic lesion);

vii. Women must be willing to comply with all requirements including the return visits and sign an informed consent.

In addition to being participants of the Ludwig-McGill cohort study, women had to satisfy the following additional criteria to be included in the sub-population for the current analysis:

viii. Must have had a negative HPV DNA detection test at the baseline-visit, in order to study only incident infections;

ix. Must have had at least one follow-up visit, with HPV testing results available;

x. Must have completed the baseline-visit interview.

All cervical cell specimens were coded in order to prevent any linking with individuals, previous HPV testing, or interview data.

2.3 Other information collected

Additional information was collected as part of the original cohort study, but not necessarily used in a formal manner for the present project. At every visit, a cervical specimen was taken for Pap cytology and blood was drawn to study the immune response to HPV. Cervicographies were also performed during the course of the study.

2.4 Data management

The data management of such a large prospective cohort study is complex. A central record database, in which the information on patient follow-up is stored, is managed by the research nurses and the epidemiologic clerk at the Ludwig Institute in São Paulo. This file contains a complete history of participation in the study, including the dates for every visit. The updated file is transferred to Montreal on a weekly basis via computer link. The data collected during the course of the study is entered into 3 different types of databases: Questionnaire Databases which contain visit-specific questionnaire information, Laboratory Results Database in which the HPV testing and serology laboratory results are filed, and the Pathology Database that records both cytology and cervicography results. The Questionnaire Databases and Laboratory Results Database are managed in Brazil. Data are entered, verified, and the updated files are transferred to Montreal. The Pathology Database is managed in Montreal, where data entry and verification for cytology and cervicography results is done by a research assistant. The 3 types of databases can be linked based on ID number according to the data needed for analysis. For this particular research project, the Questionnaire 1 Database file – questionnaire data from the baseline visit – was linked to the Laboratory Results Database file.

2.5 Status of the study

The study from which data have been extracted to conduct this research project is an ongoing process. The data analysed represent the status of this cohort as of the most recent available update of the study database, at the end of May 1997. HPV DNA detection is the limiting step of the whole data collection process, and thus, the population used for this project had to be restricted based upon availability of HPV status.

3 Statistical analysis

3.1 Descriptive statistics

The proportion of women belonging to the different category levels within each variable has been calculated for all categorical variables. The mean, median and standard deviation have been used for the continuous variables. The distribution of the study population for this research project has been compared to the interview data from the original cohort study. Where judged appropriate, equality of proportions were verified by a chi-square test and equality of means by a t-test (Armitage and Berry, 1994).

3.2 Follow-up and definition of outcome

The outcomes of interest for this research project are time to low-risk and time to high-risk HPV infection, both of which are considered as independent events. The time to an event was therefore defined as the time between the baseline visit and the visit at which HPV DNA was detected. Every woman had therefore two different outcome variables: time to low-risk and time to high-risk HPV infection of which both, one or none could be censored. Very few women, 1.2%, had both low-risk and high-risk HPV infection simultaneously detected during follow-up. These women were considered as incident cases of infection in both analyses (low-risk HPV types and high-risk HPV types). In the analysis of incidence of low-risk HPV, all women were considered at risk of developing an infection. Similarly, all women were considered at risk for a high-risk type infection; women who developed a low-risk infection during the follow-up were not excluded from this analysis. Excluding women who harboured low-risk and high-risk type infections during follow-up did not change the results.

The follow-up has the same relative starting point for all subjects, t_0, defined as the time at which the baseline visit took place. Evidently, t_0 is not the same point in real time for all subjects. Time to infection is measured from the date of the baseline visit until first detection of HPV DNA in cervical cell specimens. Subjects included in this analysis have a follow-up of either 1, 2, or 3 return visits after t_0. A large proportion of the study population is right-censored, i.e. their follow-up stops at one point when they were not infected and their event-free time is not known afterwards. Right-censoring in this particular dataset can be due to loss to follow-up, but is mainly attributed to HPV DNA status not yet available for the following visit. However, the survival analysis framework has been developed to use the complete pool of follow-up available, including that of individuals whose follow-up is right-censored. Participants' follow-up time is expressed in person-time, more specifically in women-months for the present study. Person-time is defined as "the time during which a single individual meets all the definitions for inclusion in a study, and during which any disease event occurring in the individual would be known." (Walker,1991). The person-time of observation is therefore the sum of all individual units of time spent in the study.

3.3 Incidence

The occurrence of new infections can be calculated by two methods each with its own interpretation and meaning: the cumulative incidence and the incidence density rate.

The cumulative incidence is defined as the proportion of subjects of a fixed population who have developed the health event of interest during a defined period of time. It can be viewed as a probability or a risk for developing the disease within this pre-defined time period (Rothman, 1986).

Another measure of incidence for subjects with unequal lengths of follow-up is the incidence density rate. The incidence density rate, or incidence rate was calculated by summing the number of events of interest during the reference period divided by the total pool of person-time at risk for the same period. The incidence rate is not a proportion and cannot be interpreted as the risk of developing the event of interest. It is measured in units of the reciprocal of time (Rothman, 1986). The incidence density rate calculated in this way, is the average value of the instantaneous incidence density rates, also defined as hazard rate or force of morbidity. Such an instantaneous rate is defined as (Kleinbaum, 1996):

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

where T is the time at which the event of interest takes place

t is the specific time of interest

Δt is the length of time to which the instantaneous rate applies and during which an event will be considered

The average incidence density rates for the follow-up period – of up to 4 visits with HPV status available, representing a maximum of 27 months – have been calculated for low-risk and high-risk HPV infection. Confidence intervals were calculated based on the Poisson distribution (Rothman, 1986).

3.4 Kaplan-Meier survival analysis

The Kaplan-Meier or product-limit method (Kaplan and Meier, 1958) for analysis of survival data was used to obtain the overall distribution of events, as well as the specific distribution over time for low-risk and high-risk HPV infection. It was also used for these 2 outcomes, to obtain the distribution of events over time for levels of the explanatory variables. To compare the resulting time-to-infection distributions, the log-rank test was used and two-sided p-values were generated. The log-rank statistic results from an overall comparison between 2 or more survival curves by contrasting the observed versus expected cell counts under the null hypothesis, for each time at which an event took place (Mantel, 1966). The null hypothesis (H_o) being tested consists of equivalence of all incidence curves. The log-rank statistic under the null hypothesis follows a chi-square distribution, with the degrees of freedom being equal to the variable's number of categories minus one. The SPSS statistical package version 7.5 for Windows 95 was used to perform Kaplan-Meier analysis and to obtain survival tables and log-rank statistics.

Graphical representations are showed not in terms of survival function, S(t), but in terms of its reciprocal, 1-S(t). Such graphics represent the cumulative distribution of events over time or simply stated, the incidence curves. From a clinical point of view, the cumulative probability of infection is a more interesting representation than the cumulative probability of being free of infection. The graphs were produced with the SigmaPlot package (SPSS inc.) using survival time and cumulative survival probabilities generated from the Kaplan-Meier procedure in SPSS.

3.5 Cox proportional hazards regression

3.5.1 Description of the model

In order to determine the independent effect of many variables on time to HPV infection while controlling for the effect of one another, the Cox proportional hazards regression model was used.

Developed by Cox (Cox, 1972), the proportional hazards model is a semi-parametric statistical model for analysis of survival data that allows to make use of the contribution of censored cases (Last, 1995; Kleinbaum, 1996). The Cox model is expressed mathematically in terms of the hazard function as the product of 2 quantities:

$$\mathbf{h}(t, \mathbf{X}) = \mathbf{h_o}(t) \times e^{\sum_{i=1}^{p} \beta_i X_i}$$

Where $h(t,X)$ is the hazard function at time t for an individual with the vector of explanatory variables X.

$h_o(t)$ is an arbitrary and unspecified baseline hazard function

β_i is the regression parameter associated with the i^{th} explanatory variable

X_i represents the i^{th} explanatory variable, i=1,...,p

The partial likelihood function allows to cancel out the unspecified baseline hazard $\mathbf{h_o}(t)$ and accounts for censored survival times (Cox, 1972; Cox, 1975).

The Cox proportional hazards model is of a multiplicative nature. The estimated measure of association is the hazard ratio (HR) or relative hazard, which is equal to:

$$e^{\sum_{i=1}^{p} \beta_i (X_i^* - X_i)}$$

where X_i^* is the reference category for the i^{th} explanatory variable. It indicates the increase or decrease in hazard for subjects with a particular value for an explanatory variable when compared with the reference value for that variable, while controlling for other explanatory variables entered in the model.

Two assumptions should be met for this model to be valid: independence of observations and proportionality of hazards. There is no reason to suspect that the assumption of independence between the individual observations would not be met in this particular setting. The proportional hazards assumption requires the ratio of hazards to be constant over time. We used a graphical approach in order to determine whether this latter assumption was met. For each explanatory variable, two graphs of the -ln(-ln)S(t) versus time were plotted, one for each of the 2 outcomes (time to high-risk and low-risk HPV infection) using the SPSS software (SPSS inc.,1996). Each level of the variables resulted in a graphical line, and we assessed parallelism between all graphical lines representing the levels of the variables. All variables used in this analysis met the proportional hazards assumption.

Confidence intervals for regression parameters and p-values for the Wald statistics were obtained directly from the SPSS software. We assessed statistical linear trends for ordinal variables in models that contained the explanatory variable coded in 4, 5, or 6 categories and considered as continuous instead of categorical, so that only one parameter was obtained. The p-values resulting from such models are presented as p-values for trend.

3.5.2 Model selection

We built multivariable models in a backward fashion in order to take into account the reciprocal effect of the potential explanatory variables on each other. The only variable considered as an *a priori* determinant was age, based on its importance as consistently found in prevalence studies. As suggested by Hosmer & Lemeshow (Hosmer and Lemeshow,1989), all variables with a p-value of 0.25 or less in the univariate Cox regression analysis and the ones of plausible biological importance were considered for the multivariate model. The use

of such a high p-value for selection of potential explanatory variables is based on previous work which showed that the use of more traditional levels of significance often failed to identify variables of known importance (Mickey and Greenland, 1989). Multivariable models were therefore built starting with all variables identified in such a manner. Variables were then removed one at a time based on lack of statistical significance and lack of change in parameters for variables still in the model (Hosmer and Lemeshow, 1989). The adequacy of the models was monitored by the Akaike Information Criteria (AIC) which adjusts the -2 log likelihood (-2LL) for the number of parameters in the model, penalizing when more degrees of freedom are used. Backward stepwise automated selection procedures based exclusively on the significance of the variables were also used and gave the same results.

The decision to include HPVs of unknown types in the low-risk infection category might have resulted in misclassification, since it is possible that some non-identified HPVs could be in fact high-risk types that were not detected by specific probes. In order to verify whether such misclassification took place, multivariable models were built using a restricted dataset that excluded women who had incident HPV infections of unknown type.

4 Definition of variables

We extracted a subset of variables of interest from the information collected during the baseline questionnaire-based interview. Those variables can be classified into 4 groups: 1) sociodemographic characteristics; 2) markers of sexual activity; 3) reproductive and contraceptive history; 4) lifestyle factors.

Among sociodemographic characteristics, variables considered were age, ethnicity, level of education, and family income. Markers of sexual activity were as follows: age at first sexual intercourse, number of sexual partners (lifetime, past 5 years, and past year), number of occasional partners, number of unfaithful partners (as perceived by the subject), practice of anal intercourse, and practice of cunnilingus. Reproductive and contraceptive history variables

included: years of oral contraceptive use, parity, age at menarche, use of hygienic tampons, use of non-commercial absorbents during menses, vaginal douche, genital sores, condom use, sexually-transmitted diseases, number of previous Pap tests. Lifestyle factors were comprised of only tobacco smoking and alcohol drinking. The list of variables is purposely long. Since there were no published results on determinants of incident HPV infection, the selection was exploratory although strongly based on results from prevalence studies.

The coding of each variable was carefully assessed in view of substantive knowledge from the experience of our team and empirical data distribution. The best functional form was also based on results from Kaplan-Meier incidence curves for which strong overlap between curves for successive categories was considered as an indication for merging these categories. Variable coding was also assessed experimentally in light of crude Cox proportional hazards models. Similar point estimates and extensive overlap of the 95% confidence intervals (CI) were again considered as indications for combining successive categories.

We evaluated continuous variables for linearity in their relation with HPV infection. This was done by running a Cox regression model with the continuous variables categorized into 4 or more categories and plotting the median value of these categories versus the obtained parameters (Hosmer and Lemeshow, 1989). The chosen codings are discussed in the following sections.

4.1 Sociodemographic characteristics

Age at entry in the study was obtained from the Questionnaire Database as a continuous measure. We evaluated and compared many coding schemes in order to find the best functional form for this important *a priori* confounder variable. Linearity was assessed for both low-risk and high-risk infections by categorizing age into 4 categories, quintiles, and sextiles. Results with the highest number of categories (5 and 6) did not support linearity. We therefore decided not to use age as a continuous variable. The 4-category coding (18-24, 25-34, 35-44, 45-60) resulted in a linear relationship between the parameter estimates and

medians of each category for both low-risk and high-risk infections. We chose to use this coding, but treated the 4-category age variable as an ordinal variable, resulting in only one regression parameter. This decision required that linearity be assumed not over all observations but between categories, which was supported by the linearity assessment.

Ethnicity was initially coded into 5 categories that included white, mulatto, black, Asian, and native origin. Due to the relatively small proportion of individuals in some categories, mainly Asian and native origin, it was decided to dichotomize this variable into white and non-white individuals, a coding often encountered in epidemiological studies of HPV and cervical cancer.

The level of education attained was precisely recorded, making the difference between no education at all, incomplete or complete elementary and high school, professional training or college and university degrees. However, the very small proportion of women with education higher than elementary school made it impossible to keep those distinctions. We dichotomized education into women who had at most finished elementary school and those who had at least some high school or higher training.

Family monthly income was obtained as a continuous variable, which we converted into US dollars using monthly exchange rates to correct for the heavy inflation that prevailed until mid-1994. This procedure was made necessary due to economical instability and a change of currency that took place in Brazil in 1994. Incomes reported by women during and after this period would not have been comparable to those in the previous period without a common currency. The use of the US dollar had the additional advantage of adjusting for the cost of living. Linearity assessment provided evidence for a non-linear relationship between income and time to HPV infection. Results from Kaplan-Meier and crude Cox analyses suggested no significant differences in survival times or hazard among quartiles of income. We decided to simply recode the income variable into 2 categories. The cutoff point chosen was the median value: US$265 per month.

4.2 Markers of sexual activity

Age at first intercourse was examined as a continuous variable. To assess linearity, Cox models with 4 categories (≤15, 16-17, 18-19, ≥20 yrs) were used, separately for high-risk and low-risk incident infections. Since we did not find evidence of a linear relationship between age at first intercourse and both low-risk and high-risk incident infections, this variable was categorized. The final coding was dichotomous: ≤17 years of age and 18 years or more at the time of the first sexual intercourse.

For the historically most important marker of sexual activity, number of sexual partners, 3 variables were created each referring to a different period of time: lifetime, past 5 years, past year. For lifetime exposure, the categorization was 0-1, 2-3, and 4 or more partners. We also created a binary version of the variable, with 0 or 1 partner as the reference category in comparison to 2 sexual partners or more. The dichotomous form was used for number of partners in the past 5 years and in the past year since there were very few subjects with high numbers of sexual partners.

Two new markers of sexual activity were tentatively constructed from the interview information: number of occasional partners and number of unfaithful partners. The number of occasional partners was withdrawn from the analysis to avoid collinearity, since the Pearson correlation coefficient with lifetime number of partners was 0.999. The number of unfaithful male partners, as reported by the participants, was withdrawn as well due to the low response rate to this sensitive question and to the probable unreliability of answers.

Practice of anal intercourse and cunnilingus were two other variables used as markers of sexual activity. Answers from the interview were in the form: frequently, rarely, or never. Cox models showed substantial similarity in parameter estimates and overlap in 95%CI for the frequently and rarely categories. We therefore used these 2 variables in a binary "never/ever" format.

4.3 Reproductive and contraceptive history

The length of oral contraceptive use, in years, was available from the interview information. Non-linearity of this variable was confirmed by a linearity assessment. Several coding schemes were evaluated, none of which seemed predictive of incident infections. A never/ever user scheme was finally chosen, since a coding that took into account never/shorter-time/longer-time users did not prove to be more predictive of infection while requiring estimation of an extra parameter.

Current pregnancy rather than lifetime history of pregnancy would have been an *a priori* determinant for incident HPV infection, via a decreased immunity induced by hormonal changes. Since pregnancy was an exclusion criteria for the study, current pregnancy is not documented. Total number of pregnancies was used to determine whether prolonged periods of hormonal changes would have an influence on occurrence of HPV infection. Women with no pregnancies could not be compared to women with at least one pregnancy since the former category included only 6 women (0.9%). The following coding was assessed: 0-1, 2-3, 4-5, and 6 or more pregnancies. All Kaplan-Meier curves were overlapping, suggesting no noticeable difference in survival distribution. We decided to dichotomize this variable, using 0 to 2 pregnancies as the reference category which represented approximately the first quartile of the distribution.

Age at menarche was dichotomized into early debut and expected age of menarche: up to 11 years and 12 to 19 years of age.

We coded use of hygienic tampons and of non-commercial absorbents into a binary variable. We classified women as either never or ever users. Vaginal douche was reported according to the relative frequency of use such as: always, frequently, rarely, and never. Due to the small number of women in some of the categories, we coded this variable into "never/rarely" and "always/frequently".

Some markers of genital infections or discomfort were collected as reported by the women. Vaginal or vulvar infection and genital discomfort (pain, itching, and discharge) were dichotomized into having ever/never been experienced.

The frequency of condom use was collected from the interview in the form: never, very rarely, sometimes, or always. Given the distribution of data, we decided to compare women who reported never using condoms with those who reported having ever used them.

Experience of sexually-transmitted diseases (STDs) was documented by asking women whether they had consulted a physician for such problems. Women were asked to classify the problem(s) into one (or more) of seven categories: gonorrhea, chancre, condyloma, syphilis (*Treponema pallidum*), herpes, trichomoniasis (*Trichomonas vaginalis*), and candidiasis (*Candida albicans*). We later recoded this variable into a binary format: "never/ever consulted" for STDs.

Previous screening procedures for precursors of cervical cancer were quantified by asking women how many times they had been performed on them, resulting in a continuous measure. Non-linearity of the relationship between the number of previous Pap smears and incident infections was indicated by a linearity assessment. We further decided to combine the two lower quartiles and the two upper quartiles, resulting in a binary version with a cutoff point at the median number of previous Pap cytologies: 0 to 5 and 6 or more.

4.4 Lifestyle factors

Tobacco smoking was assessed in 2 different ways: current and lifetime exposure. Current exposure was simply coded into current or not current smoker in view of the influence of tobacco by-products on the immune system. This immunological effect of tobacco was considered as a short-term one and women who were former smokers were therefore classified as being "not current" smokers. Lifetime exposure was calculated in pack-years of smoking, based on the amount and time since commencement of smoking. Non-linearity was

demonstrated and the variable was coded into 3 categories using the median number of pack-years as a cutoff point for smokers: women who never smoked, women who had smoked a maximum of 4.5 pack-years, and the ones who smoked more than 4.5 pack-years.

Alcohol drinking was available from the baseline interview in which many questions probed for the type and frequency of drinking for various alcoholic beverages. Due to the small number of women who reported drinking such beverages, the only possible contrast was "never/ever" consumed alcoholic beverages.

5 Missing values

As in any questionnaire, survey or interview, some questions were left unanswered by the study subjects. However, very few questions had missing answers. Subjects with missing values were excluded from the Kaplan-Meier analysis, as there were too few to consider as a separate category. In Cox proportional hazards regression models, two strategies were used: 1) observations with missing values were kept and coded as a separate category of the variable, and; 2) observations with missing values were simply excluded from the analysis. These 2 strategies resulted in very similar regression parameters and standard errors. The crude and age-adjusted Cox regression models reported were those obtained on all subjects with valid information for the specific variables analysed. The multivariable models were built with the sample of subjects who had no missing values for all potential explanatory variables being considered. The frequency distributions of these samples with complete information were compared to the overall sample and the incidence sample and no differences were found.

RESULTS

1 Descriptive statistics

Of the women eligible and asked to participate to the Ludwig-McGill cohort study, 70.5% agreed. A subset of this study population was selected for the present research project. Figure 2 is a flowchart representing the selection process. A total of 2528 women were enrolled in the study as of the end of March 1997. Of these women, 903 (35.7%) had HPV test results available at the analysis closing date. For this project, women who were HPV negative at the first visit were selected, reducing the subset to 746 women. When only women who had 2, 3, or 4 available HPV results were considered (1, 2 or 3 follow-up visits in addition to the baseline visit) 642 women were left for this project's study group. All were HPV negative at visit 1, and had at least 2, at most 4 HPV results available. All women selected based on HPV testing criteria had participated in the baseline interview from which information on determinants was derived. When considering low-risk HPV infection, 65 events arose from a total of 6962 women-months of follow-up. For high-risk HPV types, the follow-up totalled 6950 women-months and gave rise to 53 events. Median follow-up was 12 months with a range between 4 and 27 months.

Characteristics of the subset to be studied were compared to those of the Ludwig-McGill cohort study population. Table 4 shows the frequency distribution for selected variables measured at the baseline interview, for the subset and the whole cohort population. At the time of the analysis, the information from 1454 of the 2528 baseline interviews had been entered in the study database. The characteristics of the subset for this project were thus compared to those of the first 1454 women enrolled in the Ludwig-McGill cohort study. Only minor discrepancies were found. Frequency distributions for characteristics not presented in table 4 were also very similar for the 2 groups. The subset selected seems, therefore, to be representative of the whole study population in terms of subject characteristics.

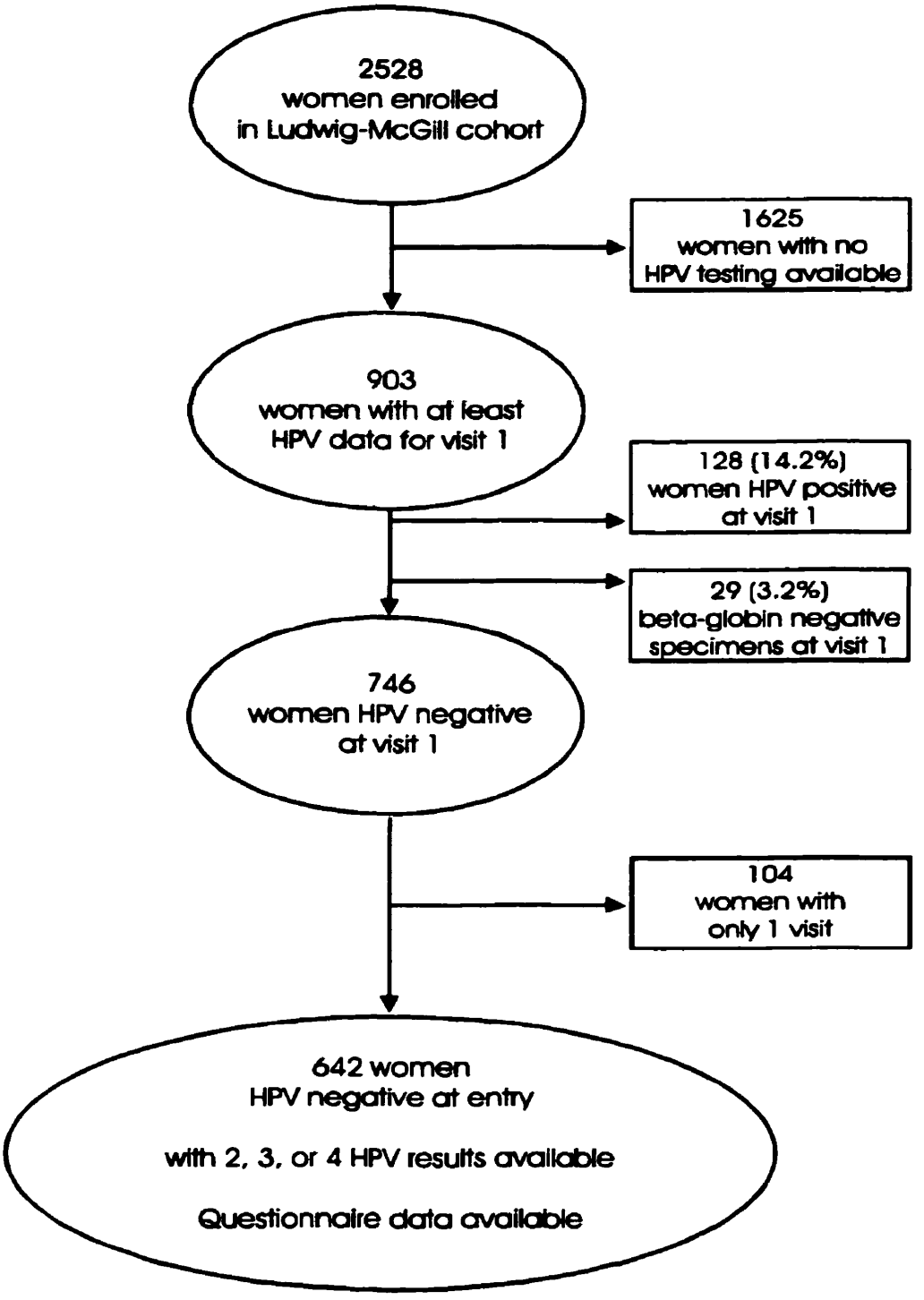


Figure 2: **Flowchart describing the selection of subjects from the Ludwig-McGill cohort study population based on availability and result of HPV testing**

Table 4: Frequency distribution of selected variables among the subset and the whole population of the Ludwig-McGill cohort study

	HPV incidence study (N=642)	Ludwig-McGill cohort study (N=1454)
Variable	**Frequency (%)**	**Frequency (%)**
Age		
18-24	108 (16.8)	289 (19.9)
25-34	258 (40.2)	584 (40.2)
35-44	203 (31.6)	408 (28.1)
45-60	73 (11.4)	173 (11.9)
Ethnicity		
White	418 (65.1)	958 (65.9)
Non-white	224 (34.5)	494 (34.0)
Missing	0	2 (0.1)
Age at first intercourse		
≤ 15	160 (24.9)	399 (27.4)
16-17	162 (25.2)	367 (25.2)
18-19	138 (21.5)	298 (20.5)
≥ 20	182 (28.3)	390 (26.8)
Lifetime # sexual partners		
0-1	299 (46.6)	643 (44.2)
2-3	210 (32.7)	491 (33.8)
≥ 4	133 (20.7)	320 (22.0)
Number of pregnancies		
0-2	232 (36.1)	548 (37.7)
≥ 3	401 (62.5)	889 (61.1)
missing	9 (1.4)	17 (1.2)
Condom use		
Never	268 (41.7)	595 (40.9)
Ever	374 (58.3)	859 (59.1)
Previous Pap smears		
0-5	340 (53.0)	791 (54.4)
≥ 6	302 (47.0)	647 (44.5)
Missing	0	16 (1.1)
Smoking status		
Not current	432 (67.3)	950 (65.3)
Current	210 (32.7)	504 (34.7)

The subset of the HPV incidence study and the Ludwig-McGill cohort were also compared in terms of follow-time, number of visits attended, and compliance to the 4-month return intervals. For the overall follow-up time without restricting based on the availability of HPV data, the HPV incidence subset had a longer median follow-up time, with 24.6 months and 5 visits as shown in table 5. This corresponds to an expected difference between the 2 groups, since the HPV incidence subset is formed mainly of the first women enrolled in the Ludwig-McGill cohort study.

Table 5: Follow-up indicators for the subset and the population of the Ludwig-McGill cohort study

	HPV incidence study (N=642)	Ludwig-McGill cohort study (N=2528)
Median follow-up time (months)	24.6	12.0
Median number of visits	5	4

The compliance to the 4-month return intervals was also compared between the HPV incidence subset and the participants of the Ludwig-McGill cohort study with more than one visit. Results shown in table 6 indicate that there were no important differences in compliance to the return intervals between the 2 groups.

Table 6: Compliance to the 4-month return intervals for the subset and the population of the Ludwig-McGill cohort study with more than one visit

	HPV incidence study (N=642)	Ludwig-McGill cohort study (N=2159)
Mean time between visit 1 & 2 (months)	4.2	4.4
Median time between visit 1 & 2 (months)	4.0	4.0
25% percentile of distribution for the time between visit 1 & 2 (months)	3.9	3.9
75% percentile of distribution for the time between visit 1 & 2 (months)	4.2	4.3

1.1 Occurrence of HPV infection and sociodemographic characteristics

Table 7 provides a sociodemographic profile of the study population. Among the 642 participating women, 461 (71.8%) were aged 25 to 44 years old. Younger and older women represented 28.2% of the study population, with 108 (16.8%) women being in the 18-24 age-group. The mean age was 33 years old and the age ranged from 18 to 59 years. In terms of ethnicity, 418 (65.1%) women were white. The study population had a low level of education: only 115 (17.9%) women had education beyond elementary school. The family monthly income ranged from US$5.00 to US$3840.00, with a median of US$265.00.

For low-risk HPV infection, the proportion of incident events did not result in statistically significant differences according to age (p=0.6346). In terms of ethnicity, white women were less likely to have developed an incident infection during follow-up, 8.6%, than non-white women, 12.9% (p=0.0827). Education, when considered as a dichotomous variable, suggested an elevated proportion of women with incident infection in the most educated group: 9.3% for women having up to elementary school compared to 13.9% for women with least some high school education (p=0.1390). A lower proportion of incident low-risk HPV infection were found in the low income group (11.5%) than in the higher (above-median) income group (8.6%), although this difference was not statistically significant (p=0.2318).

Different patterns were seen for high-risk HPV infections, for which the only variable associated with a specific distribution of incident events was age. The proportion of incident high-risk infections were 18.5%, 8.1%, 4.9%, and 2.7% respectively in the 18-24, 25-34, 35-44, and 45-60 age-groups (p=0.0001). Different categories of ethnicity, education, and income resulted in similar distributions of incident high-risk HPV infections.

Table 7: Occurrence of incident HPV infection according to sociodemographic characteristics of the study population

Variable	Subjects (N=642)	Incident HPV infections Low-risk (%)	High-risk (%)
Age			
18-24	108	9 (8.3)	20 (18.5)
25-34	258	24 (9.3)	21 (8.1)
35-44	203	22 (10.8)	10 (4.9)
45-60	73	10 (13.7)	2 (2.7)
		p[1]=0.6436	*p=0.0001*
Ethnicity			
white	418	36 (8.6)	32 (7.7)
non-white	224	29 (12.9)	21 (9.4)
		p=0.0827	*p=0.4505*
Education level			
incompl. elem.	171	15 (8.8)	14 (8.2)
elementary	355	34 (9.6)	27 (7.6)
high school	94	14 (14.9)	9 (9.6)
college/univ.	21	2 (9.5)	3 (14.3)
missing	1	0 (0.0)	0 (0.0)
		p=0.4211	*p=0.7006*
≤ elementary	526	49 (9.3)	41 (7.8)
≥ high school	115	16 (13.9)	12 (10.4)
missing	1	0 (0.0)	0 (0.0)
		p=0.1390	*p=0.3517*
Income ($US)			
0.00-157.99 (1Q)	159	14 (8.8)	15 (9.4)
158.00-264.99 (2Q)	154	22 (14.3)	15 (9.7)
265.00-419.99 (3Q)	159	12 (7.5)	13 (8.2)
420.00-3840.00 (4Q)	154	15 (9.7)	9 (5.8)
missing	16	2 (12.5)	1 (6.2)
		p=0.2176	*p=0.5912*
0.00-264.99	313	36 (11.5)	30 (9.6)
265.00-3840.00	313	27 (8.6)	22 (7.0)
missing	16	2 (12.5)	1 (6.2)
		p=0.2318	*p=0.2466*

1 Two-sided p-values for Pearson's chi-square test of equality of proportions. Missing value categories are excluded.

1.2 Occurrence of HPV infection and sexual activity

The proportion of women who have become HPV positive during follow-up, according to various markers of sexual activity, is shown in table 8. As indicated in this table, 322 (50.2%) women have had sexual intercourse at 17 years of age or earlier. Almost half of the women (299, 46.6%) had none or one lifetime sexual partner, the median being 2 partners. When the reference period was closer, respectively 115 (17.9%) and 26 (4.1%) women reported 2 or more partners for the past 5 years and the past year. Having ever experienced anal intercourse and oral sex was reported respectively by 243 (37.9%) and 341 (53.1%) women.

A slightly higher proportion of women have become positive for a low-risk type among those who had their first sexual intercourse at 18 years and older when compared to women who initiated sex at 17 years of age and younger, respectively 11.3% vs 9.0%, but these proportions are not statistically different ($p=0.3460$). A more dramatic difference is seen for high-risk type infections. Among women who had initiated sexual intercourse by the time they were 17 years old, 12.7% became positive during follow-up. However, among women who had their first intercourse at 18 years of age or older, only 3.8% became positive for a high-risk HPV type ($p<0.0001$). In terms of number of sexual partners, the highest proportion of women with incident infection is seen among those who had 2 or more sexual partners in the past year: 34.6% for both low-risk and high-risk infections. In general, subgroups of women who reported having experienced anal intercourse or oral sex gave rise to a higher proportion of incident cases. The following proportions of incident high-risk infections were obtained for women who never had anal intercourse and those who reported having had at least one: 6.8% and 10.7% respectively ($p=0.0790$). The proportions of incident low-risk infections were 8.0% and 12.0% respectively in women reporting never and ever having experienced oral sex ($p=0.0896$).

Table 8: Occurrence of incident HPV infection according to markers of sexual activity

Variable	Subjects (N=642)	Incident HPV infections Low-risk (%)	High-risk (%)
Age at first intercourse			
≤ 15 years	160	15 (9.4)	23 (14.4)
16-17	162	14 (8.6)	18 (11.1)
18-19	138	15 (10.9)	8 (5.8)
≥ 20 years	182	21 (11.5)	4 (2.2)
		p[1]=0.8075	*p=0.0002*
≤ 17 years	322	29 (9.0)	41 (12.7)
≥ 18 years	320	36 (11.3)	12 (3.8)
		p=0.3460	*p<0.0001*
Lifetime # partners			
0-1	299	25 (8.4)	15 (5.0)
2-3	210	14 (6.7)	24 (11.4)
≥ 4	133	26 (19.5)	14 (10.5)
		p=0.0002	*p=0.0199*
0-1	299	25 (8.4)	15 (5.0)
≥ 2	343	40 (11.7)	38 (11.1)
		p=0.1667	*p=0.0054*
# partners past 5 years			
0-1	527	39 (7.4)	34 (6.5)
≥ 2	115	26 (22.6)	19 (16.5)
		p<0.0001	*p=0.0004*
# partners past year			
0-1	607	56 (9.2)	44 (7.2)
≥ 2	26	9 (34.6)	9 (34.6)
missing	9	0 (0.0)	0 (0.0)
		p=0.0006	*p=0.0001*
Anal intercourse			
Never	399	43 (10.8)	27 (6.8)
Ever	243	22 (9.1)	26 (10.7)
		p=0.4826	*p=0.0790*
Oral sex			
Never	301	24 (8.0)	21 (7.0)
Ever	341	41 (12.0)	32 (9.4)
		p=0.0896	*p=0.2687*

1 Two-sided p-values for Pearson's chi-square test of equality of proportions. Missing value categories are excluded.

1.3 Occurrence of HPV infection according to reproductive and contraceptive history

Table 9 shows the reproductive and contraceptive profile of the study group, as well as the occurrence of incident infection for each level of the variables. A group of 101 (15.7%) women reported having never used OCs as a contraceptive method. The 541 women who had used OCs could be divided into 349 (64.5%) who had used OCs for less than 6 years and 192 (35.5%) who reported having used OCs for 6 years or more. The study population included 87 (13.6%) women having had 0 or 1 pregnancy, 292 (45.5%) women reporting 2 or 3 pregnancies, 149 (23.2%) women reporting 4 or 5 pregnancies, and 105 (16.4%) participants reporting 6 or more pregnancies. The age at menarche was from 8 to 11 years of age for 137 (21.3%) women and 12 to 19 years of age for 504 (78.5%) women. Most women, 570 (88.8%) of them, reported never using hygienic tampons, while 278 (43.3%) subjects reported having used non-commercial absorbents during their menses. Vaginal douching was reported as a frequent practice by 65 (10.1%) participants. Vaginal or vulvar infection had been experienced by 99 (15.4%) women. Genital discomfort, defined as pain, itching or discharge, had been experienced in the past 5 years by 253 (39.4%) women and in the past 2 days by 228 (35.5%) women. A group of 268 (4.0%) participants reported never using the condom, whereas 232 (36.1%) reported using condoms rarely, and 142 (22.1%) reported using them frequently. Physician consultations for STDs were reported by 121 (18.8%) women. A little over half of the women, 340 (53.0%), reported having had 0 to 5 previous Pap cytologies.

For low-risk infections, no important difference in distribution of incident events were seen among the levels of the explanatory variables except for vaginal douching and consultation for STDs. Out of the women reporting "never or rarely" using vaginal douches compared to those who reported "always or frequently" using them, 10.7% and 4.6% respectively developed low-risk HPV infection during follow-up ($p=0.1204$). Among women who consulted a physician for STDs, 14.0% developed a low-risk infection compared to 9.3% among those who reported no such consultations ($p=0.1188$).

Regarding high-risk infections, the number of pregnancies, condom use, and previous Pap tests were associated with proportions of incident events. Women with fewer pregnancies (0 to 2) comprised a higher proportion of incident events, 10.3%, than women with 3 or more pregnancies, 7.0% (p=0.1377). Among the group of women who reported never using condoms, 10.4% developed high-risk infection during follow-up compared to 6.7% of the women who reported having ever used condoms (p=0.0875). Among women who had had a maximum of 5 previous Pap smears, 10.9% developed a high-risk infection whereas only 5.3% of the women who reported having 6 or more previous Pap smears did so (p=0.0103).

1.4 Occurrence of HPV infection according to lifestyle factors

Table 10 shows the population distribution according to tobacco smoking and alcohol drinking. One-third (32.7%) of the subjects were current smokers. Lifetime exposure to tobacco smoking ranged from 0 to 55.55 pack-years, the median being 4.50 pack-years. Approximately one-third (31.0%) of the women reported never drinking alcohol.

Neither lifetime smoking exposure, nor alcohol drinking seemed to be associated with incident low-risk or high-risk HPV infection. However, smoking status resulted in different proportions of incident events among current and non-current smokers. Among women who did not smoke at the time of the baseline interview, 8.8% developed a low-risk infection during follow-up compared to 12.9% among current smokers (p=0.1095). For high-risk infections, the proportions of incident events were 6.9% and 11.0% respectively among non-current and current smokers (p=0.0834).

Table 9: Occurrence of incident HPV infection according to reproductive and contraceptive history

Variable	Subjects (N=642)	Incident HPV infections Low-risk (%)	High-risk (%)
Years of OC use			
0.00-0.25 (1Q)	160	15 (9.4)	16 (10.0)
0.26-3.00 (2Q)	152	13 (8.6)	11 (7.2)
3.01-7.00 (3Q)	162	21 (13.0)	14 (8.6)
7.01-29.00 (4Q)	168	16 (9.5)	12 (7.1)
		p'=0.5698	*p=0.7632*
0	101	9 (8.9)	10 (9.9)
< 6 years	349	37 (10.6)	26 (7.4)
≥ 6 years	192	19 (9.9)	17 (8.9)
		p=0.8773	*p=0.6869*
Never	101	9 (8.9)	10 (9.9)
Ever	541	56 (10.4)	43 (7.9)
		p=0.6596	*p=0.5127*
Number of pregnancies			
0-1	87	10 (11.5)	8 (9.2)
2-3	292	29 (9.9)	26 (8.9)
4-5	149	12 (8.1)	10 (6.7)
≥ 6	105	13 (12.4)	8 (7.6)
missing	9	1 (11.1)	1 (11.1)
		p=0.6863	*p=0.8516*
0-2	232	24 (10.3)	24 (10.3)
≥ 3	401	40 (10.0)	28 (7.0)
missing	9	1 (11.1)	1 (11.1)
		p=0.8818	*p=0.1377*
Age at menarche			
8-11	137	15 (10.9)	15 (10.9)
12-19	504	50 (9.9)	38 (7.5)
missing	1	—	—
		p=0.7237	*p=0.1989*
Use of hygienic tampons			
Never	570	60 (10.5)	46 (8.1)
Ever	72	5 (6.9)	7 (9.7)
		p=0.3424	*p=0.6313*
Use of non-commercial absorbents			
Never	364	32 (8.8)	33 (9.1)
Ever	278	33 (11.9)	20 (7.2)
		p=0.2000	*p=0.3932*

Table 9: Occurrence of incident HPV infection according to reproductive and contraceptive history (continued)

Variable	Subjects (N=642)	Incident HPV infections Low-risk (%)	High-risk (%)
Vaginal douching			
Never/rarely	577	62 (10.7)	48 (8.3)
Always/frequently	65	3 (4.6)	5 (7.7)
		p=0.1204	*p=0.8619*
Vaginal/vulvar infection			
Never	543	54 (9.9)	44 (8.1)
Ever	99	11 (11.1)	9 (9.1)
		p=0.7235	*p=0.7426*
Genital discomfort past 5 yrs			
No	389	40 (10.3)	37 (9.5)
Yes	253	25 (9.9)	16 (6.3)
		p=0.8692	*p=0.1516*
Genital discomfort past 2 ds			
No	414	47 (11.4)	36 (8.7)
Yes	228	18 (7.9)	17 (7.5)
		p=0.1645	*p=0.5850*
Condom use			
Never	268	30 (11.2)	28 (10.4)
Rarely	232	26 (11.2)	14 (6.0)
Frequently	142	9 (6.3)	11 (7.7)
		p=0.2378	*p=0.1959*
Never	268	30 (11.2)	28 (10.4)
Ever	374	35 (9.4)	25 (6.7)
		p=0.4470	*p=0.0875*
Consultation for STDs			
Never	517	48 (9.3)	41 (7.9)
Ever	121	17 (14.0)	12 (9.9)
missing	4	—	—
		p=0.1188	*p=0.4759*
Previous Pap cytologies			
0-2 (1Q)	154	12 (7.8)	19 (12.3)
3-5 (2Q)	186	20 (10.8)	18 (9.7)
6-10 (3Q)	149	19 (12.8)	7 (4.7)
≥ 11 (4Q)	153	14 (9.2)	9 (5.9)
		p=0.5142	*p=0.0572*
0-5	340	32 (9.4)	37 (10.9)
≥ 6	302	33 (10.9)	16 (5.3)
		p=0.5252	*p=0.0103*

1 Two-sided p-values for Pearson's chi-square test of equality of proportions. Missing value categories are excluded.

Table 10: Occurrence of incident HPV infection according to lifestyle factors

Variable	Subjects (N=642)	Incident HPV infections Low-risk (%)	High-risk (%)
Smoking status			
Not current smoker	432	38 (8.8)	30 (6.9)
Current smoker	210	27 (12.9)	23 (11.0)
		p[1]=0.1095	*p=0.0834*
Lifetime smoking exposure (pack-years)			
0	324	29 (9.0)	24 (7.4)
0.01-1.40 (1Q)	77	6 (7.8)	7 (9.1)
1.41-5.40 (2Q)	79	11 (13.9)	6 (7.6)
5.41-11.20 (3Q)	71	9 (12.7)	7 (9.9)
11.21-55.55 (4Q)	91	10 (11.0)	9 (9.9)
		p=0.5941	*p=0.9119*
0	324	29 (9.0)	24 (7.4)
0.01-4.50	148	17 (11.5)	12 (8.1)
4.51-55.55	170	19 (11.2)	17 (10.0)
		p=0.6068	*p=0.6080*
Alcohol drinking			
Never	199	16 (8.0)	14 (7.0)
Ever	442	49 (11.1)	39 (8.8)
missing	1	—	—
		p=0.2372	*p=0.4469*

1 Two-sided p-values for Pearson's chi-square test of equality of proportions. Missing value categories are excluded.

2 Incidence density rates of HPV infection

The follow-up period ranged from 4 to 27 months, with a median of 12 months. Incidence density rates have been calculated for low-risk and high-risk HPV infection, as two independent events. An overall incidence density rate has also been calculated, regardless of HPV type or classification. Table 11 shows the incidence rates and their respective confidence intervals for low-risk and high-risk infections, and for specific age-groups.

Table 11: Incidence rate of HPV infection by risk and age

HPV infection	Age	Events	Follow-up (women-months)	Incidence rate (/1000 women-months)	95% CI
Overall		**110**	**6711**	**16.4**	**(13.3 - 19.4)**
Low-risk		**65**	**6962**	**9.3**	**(7.1 - 11.6)**
	18-24	9	1158	7.8	(2.7 - 12.8)
	25-34	24	2857	8.4	(5.0 - 11.8)
	35-44	22	2188	10.0	(5.8 - 14.2)
	45-60	10	758	13.2	(5.0 - 21.4)
				P-value for trend = 0.196	
High-risk		**53**	**6950**	**7.6**	**(5.6 - 9.7)**
	18-24	20	1058	18.9	(10.6 - 27.2)
	25-34	21	2875	7.3	(4.2 - 10.4)
	35-44	10	2212	4.5	(1.7 - 7.3)
	45-60	2	804	2.5	(0.0 - 5.9)
				P-value for trend = 2.4×10^{-5}	

Differences can be noted in the distribution of incident infections according to classification of HPV type (low-risk or high-risk) for the different age-groups. There was a slight increase in incidence rate with increasing age for low-risk HPV types, but the p-value for trend indicated no statistically important differences (trend p-value=0.196). The pattern seen with high-risk infection was different. A marked decrease of incidence rates was observed with increasing age. The p-value obtained when a test for trend was performed for the age-specific incidence rates of high-risk HPV infection was highly significant (trend p-value=2.4×10^{-5}).

Using a previous version of the dataset, the overall incidence rates of HPV infection were calculated while stratifying for the number of return visits, in order to determine whether the frequency of testing had an effect on the estimated incidence rates. These incidence rates are presented in table 12. Women included in this analysis, as in all others presented were HPV negative at the baseline visit.

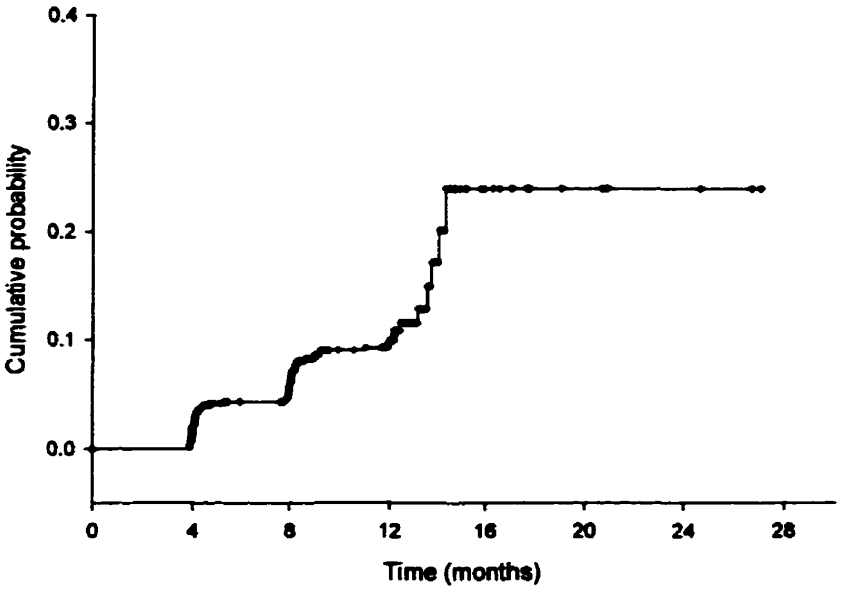
Table 12: Effect of testing frequency on incidence rates of overall HPV infection

Return visits	Events	Follow-up (women-months)	Incidence rate (/1000 women-months)	95% CI
1	5	342	14.6	(1.8 - 27.4)
2	26	1651	15.7	(9.7 - 21.8)
3	56	3692	15.2	(11.2 - 19.1)

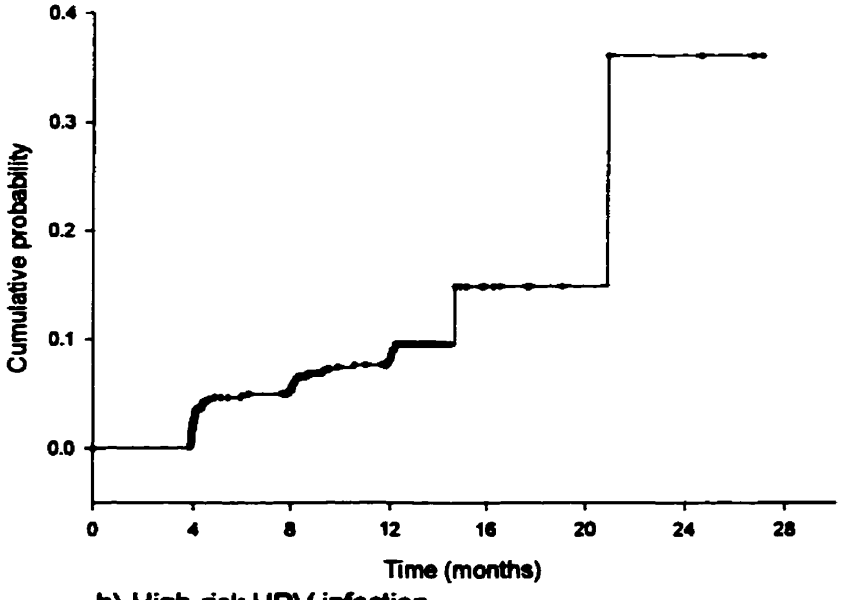
Except for the relative imprecision of the incidence rate for women who had only one visit, due to small numbers, the incidence rates do seem stable within all three return visits categories. The cumulative number of events and therefore cumulative incidence were artificially increased by the frequency of HPV testing, but the incidence rates were not.

3 Kaplan-Meier analysis

The cumulative distribution of newly detected HPV infection, 1-S(t), is presented in figure 3. Separate incidence curves are presented for high-risk and low-risk HPV infection. The cumulative probability and standard error (SE) of developing a low-risk HPV infection during the follow-up was 0.043 (0.008) at 6 months and 0.097 (0.012) at 12 months. For development of a high-risk infection, the cumulative probability (SE) was 0.046 (0.008) at 6 months and 0.080 (0.011) at 12 months.



a) Low-risk HPV infection



b) High-risk HPV infection

Figure 3: **Incidence curves**

Cumulative distribution of events [1-S(t)] over the follow-up period for a) low-risk HPV infection , and b) high-risk HPV infection.

3.1 Kaplan-Meier analysis according to sociodemographic characteristics

Table 13 presents the cumulative rates of HPV infection for low- and high-risk types according to sociodemographic characteristics. Those rates have been calculated as the proportion of events occurred at respectively 6 and 12 months of follow-up in the risk set, and are used to summarize the 1-S(t) curves. For low-risk HPV infection, different ages, education levels, and incomes did not result in significantly different distribution of events. In terms of ethnicity, non-white women had cumulative rates of low-risk infection close to twice as high as those of white women at both 6 and 12 months ($p=0.085$ log-rank test[1]). For high-risk HPV infection, the only sociodemographic factor that resulted in statistically different distribution of events was age. The highest cumulative rates were seen in the younger age-group, 18 to 24 years old, with the cumulative probabilities at 6 months and 12 months being respectively 13.5% and 19%. In the 25 to 34 years age-group, the cumulative probabilities were down to 4.6% and 5.2% respectively at 6 months and 12 months.

3.2 Kaplan-Meier analysis according to markers of sexual activity

As shown in table 14, significantly higher probabilities for a low-risk infection were seen in women with 2 or more partners in the past 5 years and in the past year when compared to the reference category ($p<0.001$ log-rank test for both variables). Oral sex also resulted in higher cumulative rates at 6 and 12 months in women who reported having experienced it ($p=0.082$ for log-rank test). All markers of sexual activity seemed associated with the distributions of high-risk infections. Younger age at first intercourse, higher number of sexual partners, and practice of anal intercourse all resulted in higher cumulative rates than the reference category. Practice of oral sex also resulted in elevated cumulative rates, but was not statistically significant ($p=0.172$ log-rank test).

---

[1] P-value for the log-rank test for the overall comparison among the incidence curves

Table 13: Cumulative rates of HPV infection according to sociodemographic characteristics

	Cumulative risk of HPV infection					
	Low-risk HPV types			High-risk HPV types		
Variable	**6 mos (SE)**	**12 mos (SE)**	**p[1]**	**6 mos (SE)**	**12 mos (SE)**	**p**
Age						
18-24	0.029 (0.016)	0.086 (0.029)	0.519	0.135 (0.034)	0.190 (0.040)	0.001
25-34	0.040 (0.012)	0.084 (0.018)		0.024 (0.010)	0.068 (0.017)	
35-44	0.045 (0.015)	0.105 (0.023)		0.046 (0.015)	0.052 (0.016)	
45-60	0.071 (0.031)	0.136 (0.042)		0.000 (0.000)	0.033 (0.023)	
Ethnicity						
White	0.035 (0.009)	0.077 (0.014)	0.085	0.034 (0.009)	0.069 (0.013)	0.440
Non-white	0.060 (0.016)	0.134 (0.024)		0.069 (0.017)	0.100 (0.021)	
Education level						
≤ elementary	0.041 (0.009)	0.089 (0.013)	0.143	0.049 (0.010)	0.074 (0.012)	0.356
≥ high school	0.053 (0.021)	0.132 (0.033)		0.036 (0.018)	0.108 (0.031)	
Income (US$)						
0-264.99	0.036 (0.011)	0.103 (0.018)	0.276	0.056 (0.013)	0.090 (0.017)	0.259
265-3840	0.053 (0.013)	0.087 (0.017)		0.036 (0.011)	0.070 (0.015)	

1 Two-sided p-value for the log-rank test of equality among the survival distributions.

Table 14: Cumulative rates of HPV infection according to markers of sexual activity

	Cumulative risk of HPV infection					
	Low-risk HPV types			High-risk HPV types		
Variable	6 mos (SE)	12 mos (SE)	p[1]	6 mos (SE)	12 mos (SE)	p
Age 1st interc.						
≤ 17 years	0.039 (0.011)	0.088 (0.017)	0.284	0.074 (0.015)	0.126 (0.020)	0.001
≥ 18 years	0.048 (0.012)	0.105 (0.018)		0.019 (0.008)	0.034 (0.010)	
Lifetime # part.						
0-1	0.038 (0.011)	0.080 (0.016)	0.180	0.034 (0.011)	0.042 (0.012)	0.006
≥ 2	0.049 (0.012)	0.112 (0.018)		0.058 (0.013)	0.114 (0.018)	
# part. past 5 yrs						
0-1	0.029 (0.008)	0.072 (0.012)	0.001	0.037 (0.008)	0.057 (0.011)	0.001
≥ 2	0.108 (0.030)	0.208 (0.040)		0.089 (0.027)	0.183 (0.038)	
# part. past year						
0-1	0.041 (0.008)	0.089 (0.012)	0.001	0.042 (0.008)	0.068 (0.011)	0.001
≥ 2	0.119 (0.064)	0.303 (0.097)		0.158 (0.072)	0.403 (0.107)	
Anal intercourse						
Never	0.039 (0.010)	0.101 (0.016)	0.520	0.038 (0.010)	0.065 (0.013)	0.054
Ever	0.051 (0.014)	0.089 (0.019)		0.060 (0.016)	0.105 (0.021)	
Oral sex						
Never	0.034 (0.011)	0.081 (0.017)	0.082	0.034 (0.010)	0.061 (0.014)	0.172
Ever	0.052 (0.012)	0.110 (0.018)		0.058 (0.013)	0.097 (0.017)	

1 Two-sided p-value for the log-rank test of equality among the survival distributions.

3.3 Kaplan-Meier analysis according to reproductive and contraceptive history

Only a few among the many variables from this group resulted in different distribution of events during follow-up. For low-risk HPV types, these factors were: use of non-commercial absorbents, vaginal douching, and consultation for STDs. Women who reported having used non-commercial absorbents, i.e. reusable cloths, had cumulative probabilities of developing a low-risk infection of 4.8% and 12.2% compared to 4.0% and 7.8% for women who did not, respectively at 6 and 12 months (p=0.116, log-rank test). Frequent vaginal douching was associated with lower cumulative rates, with 1.7% and 5.3%, compared to rare or no douching, with 4.6% and 10.2%, at 6 and 12 months respectively (p=0.109, log-rank test). Consultation for STDs resulted in higher cumulative rates, 6.8% and 13.3%, than having never consulted for such matter, 3.8% and 8.9%, at 6 and 12 months (p=0.069, log-rank test).

Two other variables seemed associated with the distribution of events for high-risk infections: condom use and the number of previous Pap smears. Women who reported having ever used condoms had lower cumulative rates of high-risk HPV infection than women who reported never using condoms: 4.1% and 6.0% compared to 5.4% and 11.0% respectively at 6 and 12 months (p=0.097, log-rank test). Women who reported having had 6 or more previous Pap smears had much lower cumulative rates of high-risk infection than women who reported 0 to 5 previous Pap smears: 2.0% and 4.8% compared to 7.0% and 10.9% respectively at 6 and 12 months (p=0.008).

3.4 Kaplan-Meier analysis according to lifestyle factors

For both low- and high-risk HPV infection, current smokers had higher cumulative rates of infection at 6 and 12 months (p=0.096, log-rank test for low-risk HPV; p=0.100, log-rank test for high-risk HPV). Women who reported ever drinking alcohol had also slightly elevated cumulative rates of HPV infection.

Table 15: Cumulative rates of HPV infection according to reproductive and contraceptive history

	Cumulative risk of HPV infection					
	Low-risk HPV types			High-risk HPV types		
Variable	6 mos (SE)	12 mos (SE)	p[1]	6 mos (SE)	12 mos (SE)	p
OC use						
Never	0.022 (0.016)	0.087 (0.032)	0.866	0.088 (0.030)	0.114 (0.034)	0.359
Ever	0.047 (0.009)	0.099 (0.013)		0.039 (0.008)	0.074 (0.012)	
# pregnancies						
0-2	0.036 (0.012)	0.092 (0.020)	0.981	0.053 (0.015)	0.099 (0.021)	0.154
≥ 3	0.046 (0.011)	0.098 (0.016)		0.041 (0.010)	0.068 (0.013)	
Age menarche						
8-11	0.067 (0.022)	0.117 (0.029)	0.667	0.067 (0.022)	0.110 (0.028)	0.219
12-19	0.037 (0.009)	0.091 (0.014)		0.041 (0.009)	0.072 (0.012)	
Use of tampons						
Never	0.044 (0.009)	0.101 (0.013)	0.375	0.050 (0.009)	0.080 (0.012)	0.615
Ever	0.043 (0.024)	0.061 (0.030)		0.015 (0.021)	0.083 (0.036)	
Non-comm. abs.						
Never	0.040 (0.010)	0.078 (0.015)	0.116	0.040 (0.010)	0.084 (0.015)	0.477
Ever	0.048 (0.013)	0.122 (0.021)		0.055 (0.014)	0.074 (0.017)	
Vaginal douche						
Never/rare	0.046 (0.009)	0.102 (0.013)	0.109	0.048 (0.009)	0.079 (0.012)	0.799
Always/freq.	0.017 (0.017)	0.053 (0.030)		0.031 (0.022)	0.089 (0.038)	
Vag/vulv infect.						
Never	0.046 (0.009)	0.095 (0.013)	0.640	0.043 (0.009)	0.076 (0.012)	0.624
Ever	0.032 (0.018)	0.104 (0.033)		0.064 (0.025)	0.100 (0.032)	
Discomfort 5Y						
No	0.042 (0.010)	0.093 (0.016)	0.952	0.053 (0.011)	0.093 (0.016)	0.184
Yes	0.046 (0.013)	0.102 (0.020)		0.037 (0.012)	0.060 (0.016)	
Discomfort 2D						
No	0.045 (0.010)	0.104 (0.016)	0.274	0.044 (0.010)	0.082 (0.014)	0.717
Yes	0.041 (0.013)	0.083 (0.020)		0.050 (0.015)	0.076 (0.018)	
Condom use						
Never	0.043 (0.013)	0.105 (0.020)	0.432	0.054 (0.014)	0.110 (0.020)	0.097
Ever	0.044 (0.011)	0.091 (0.016)		0.041 (0.010)	0.060 (0.013)	
STDs						
Never	0.038 (0.009)	0.089 (0.013)	0.069	0.044 (0.009)	0.078 (0.012)	0.350
Ever	0.068 (0.023)	0.133 (0.033)		0.060 (0.022)	0.092 (0.028)	
# Previous Paps						
0-5	0.049 (0.012)	0.099 (0.017)	0.519	0.070 (0.014)	0.109 (0.018)	0.008
≥ 6	0.037 (0.011)	0.094 (0.018)		0.020 (0.008)	0.048 (0.013)	

1 Two-sided p-value for the log-rank test of equality among the survival distributions.

Table 16: Cumulative rates of HPV infection according to lifestyle factors

	Cumulative risk of HPV infection					
	Low-risk HPV types			High-risk HPV types		
Variable	6 mos (SE)	12 mos (SE)	p[1]	6 mos (SE)	12 mos (SE)	p
Smoking						
Not current	0.036 (0.009)	0.082 (0.014)	0.096	0.036 (0.009)	0.069 (0.013)	0.100
Current	0.058 (0.016)	0.126 (0.024)		0.068 (0.018)	0.102 (0.022)	
Alcohol drinking						
Never	0.031 (0.012)	0.077 (0.020)	0.222	0.036 (0.013)	0.054 (0.017)	0.422
Ever	0.049 (0.010)	0.106 (0.016)		0.052 (0.011)	0.092 (0.014)	

1 Two-sided p-value for the log-rank test of equality among the survival distributions.

4 Cox proportional hazards regression

After assessing the proportional hazards assumption with -ln(-ln S(t)) graphs, Cox proportional hazards regression analysis were computed for the 2 outcomes of interest, time to low-risk and high-risk HPV infections. Tables 17 to 20 show the results of crude and age-adjusted Cox proportional hazards regression models for the 4 families of explanatory variables. The age-adjustments were done with age divided into 4 categories (18-24, 25-34, 35-44, 45-60), treated as an ordinal variable in the models. This form assumed linearity for the 4 levels, which has been verified graphically for low-risk and high-risk infection. In addition, using this form requires estimation of only one parameter instead of three in the Cox models.

4.1 Crude and age-adjusted Cox models for sociodemographic characteristics

The association found between incident low-risk infections and age indicated an increasing hazard of infections with increasing age, although not statistically significant (table 17). Non-white women also had a higher risk of developing low-risk infections, the hazard ratio (HR) being 1.53 (95%CI: 0.94-2.50) when compared to white women. This association remained unchanged after adjustment for age. A positive association was also found between the level of education and incident low-risk infection, and increased after adjustment for age. The hazard of developing a low-risk infection during follow-up for women with at least some high school education was 1.66 (95%CI: 0.93-2.95) times that of women who reported at most elementary school.

Age was strongly associated with incident high-risk infection. However, the association was in the opposite direction as the one seen for low-risk HPV, older age being associated with a lower hazard of high-risk infection. Comparatively to HRs obtained for low-risk infections, the ones for high-risk all excluded one and were strongly significant: 0.38 (95%CI: 0.20-0.70) for women 25 to 34 years, 0.23 (95%CI: 0.10-0.49) for women 35 to 44 years, and 0.12 (95%CI: 0.03-0.53) for women 45 to 60 years, when compared to women 18 to 24 years of age as the reference category (p-value for trend <0.0001).

4.2 Crude and age-adjusted Cox models for markers of sexual activity

Table 18 shows the associations between various markers of sexual activity and incident low-risk and high-risk HPV infection. For low-risk infection, associations were seen with number of sexual partners, with the more recent sexual history such as past 5 years and past year resulting in associations of the strongest magnitude. The age-adjusted hazard of low-risk infection was 4.51 (95%CI: 2.21-9.21) times higher in women who had 2 or more sexual partners in the past year compared to women who had 0 or 1 sexual partner. Women who reported having ever had oral sex performed on them were also more likely to have developed low-risk HPV infections, age-adjusted HR=1.69 (95%CI: 1.01-2.83) when compared to women who reported having never had oral sex performed on them.

High-risk infections were associated with all markers of sexual activity, although with borderline significance for anal intercourse, and not statistically significant with oral sex. Older age at first intercourse was associated with a lower hazard of high-risk infection, age-adjusted HR=0.38 (95%CI: 0.20-0.74) for women who had 18 years or more compared to women who had 17 years or less at the time of their first sexual intercourse. Statistically significant positive associations were obtained for numbers of sexual partners for all reference periods, with the strongest association seen for the more recent sexual history, the past year. The hazard for high-risk HPV infection during follow-up among women who reported 2 or more sexual partners in the past year was 4.89 (95%CI: 2.37-10.08) times higher than for women who had 0 or 1 partner. Having ever experienced anal intercourse was associated with a higher risk: age-adjusted HR=1.64 (95%CI: 0.95-2.83) for women who ever experienced anal intercourse compared to women who never did.

Table 17: Crude and age-adjusted Cox proportional hazards regression models for sociodemographic characteristics

	Low-risk incident HPV infection				High-risk incident HPV infection			
	Crude		Age-adjusted		Crude		Age-adjusted	
Variables	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**
Age								
18-24	1.00	*p[1]=0.1701*			1.00	*p<0.0001*		
25-34	1.07	(0.50-2.30)			0.38	(0.20-0.70)		
35-44	1.28	(0.59-2.79)			0.23	(0.10-0.49)		
45-60	1.77	(0.72-4.37)			0.12	(0.03-0.53)		
18-24	1.00				1.00			
25-60	1.24	(0.61-2.50)			0.29	(0.16-0.50)		
Ethnicity								
white	1.00		1.00		1.00		1.00	
non-white	1.53	(0.94-2.50)	1.52	(0.93-2.48)	1.24	(0.72-2.15)	1.31	(0.75-2.27)
Education level								
incompl. elementary	1.00	*p=0.3316*	1.00	*p=0.2147*	1.00	*p=0.4504*	1.00	*p=0.9832*
elementary	1.03	(0.56-1.89)	1.10	(0.60-2.04)	0.94	(0.49-1.80)	0.76	(0.40-1.46)
high school	1.72	(0.83-3.56)	2.04	(0.96-4.34)	1.20	(0.52-2.78)	0.80	(0.34-1.88)
college/university	0.92	(0.21-4.01)	0.97	(0.22-4.26)	1.72	(0.49-5.99)	1.44	(0.41-5.02)
≤ elementary	1.00		1.00		1.00		1.00	
≥ high school	1.52	(0.86-2.67)	1.66	(0.93-2.95)	1.35	(0.71-2.58)	1.09	(0.57-2.09)
Income (US$)								
0.00-157.99 (1Q)	1.00	*p=0.8185*	1.00	*p=0.7501*	1.00	*p=0.2305*	1.00	*p=0.3579*
158.00-264.99 (2Q)	1.67	(0.85-3.26)	1.67	(0.85-3.26)	1.03	(0.50-2.10)	0.98	(0.48-2.01)
265.00-419.99 (3Q)	0.86	(0.40-1.86)	0.84	(0.39-1.81)	0.86	(0.41-1.80)	0.95	(0.45-2.00)
420.00-3840.00 (4Q)	1.16	(0.56-2.41)	1.13	(0.55-2.36)	0.62	(0.27-1.41)	0.66	(0.29-1.51)
0.00-264.99	1.00		1.00		1.00		1.00	
265.00-3840.00	0.76	(0.46-1.25)	0.74	(0.45-1.22)	0.73	(0.42-1.26)	0.81	(0.47-1.41)

1 P-value for trend for variables with 3 or more categories, obtained by treating the categorical variable as ordinal in a Cox regression model.

Table 18: Crude and age-adjusted Cox proportional hazards regression models for markers of sexual activity

	Low-risk incident HPV infection				High-risk incident HPV infection			
	Crude		Age-adjusted		Crude		Age-adjusted	
Variables	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**
Age at first intercourse								
≤ 15 years	1.00	*p[1]=0.3730*	1.00	*p=0.6356*	1.00	*p<0.0001*	1.00	*p=0.012*
16-17 years	0.84	(0.40-1.74)	0.84	(0.40-1.75)	0.71	(0.38-1.32)	0.70	(0.38-1.30)
18-19 years	1.14	(0.56-2.34)	1.10	(0.54-2.26)	0.38	(0.17-0.85)	0.42	(0.19-0.94)
≥ 20 years	1.23	(0.63-2.39)	1.10	(0.55-2.20)	0.14	(0.05-0.42)	0.22	(0.07-0.65)
≤ 17 years	1.00		1.00		1.00		1.00	
≥ 18 years	1.31	(0.80-2.13)	1.20	(0.72-2.01)	0.29	(0.15-0.55)	0.38	(0.20-0.74)
Lifetime number of partners								
0-1	1.00	*p=0.0034*	1.00	*p=0.0036*	1.00	*p=0.0241*	1.00	*p=0.0036*
2-3	0.78	(0.41-1.51)	0.78	(0.40-1.50)	2.33	(1.22-4.46)	2.36	(1.23-4.51)
≥ 4	2.45	(1.41-4.24)	2.44	(1.41-4.22)	2.13	(1.03-4.42)	2.34	(1.13-4.87)
0-1	1.00		1.00		1.00		1.00	
≥ 2	1.41	(0.85-2.32)	1.40	(0.85-2.31)	2.25	(1.24-4.10)	2.35	(1.29-4.29)
Number of partners past 5 yrs								
0-1	1.00		1.00		1.00		1.00	
≥ 2	3.17	(1.93-5.21)	3.53	(2.12-5.86)	2.63	(1.50-4.61)	2.10	(1.19-3.72)
Number of partners past year								
0-1	1.00		1.00		1.00		1.00	
≥ 2	4.20	(2.07-8.51)	4.51	(2.21-9.21)	5.67	(2.76-11.66)	4.89	(2.37-10.08)
Anal intercourse								
never	1.00		1.00		1.00		1.00	
ever	0.85	(0.51-1.41)	0.87	(0.52-1.46)	1.69	(0.98-2.92)	1.64	(0.95-2.83)
Oral sex								
never	1.00		1.00		1.00		1.00	
ever	1.56	(0.94-2.58)	1.69	(1.01-2.83)	1.47	(0.84-2.58)	1.22	(0.70-2.15)

1 P-value for trend for variables with 3 or more categories, obtained by treating the categorical variable as ordinal in a Cox regression model.

4.3 Crude and age-adjusted Cox models for reproductive and contraceptive history

In crude models, only one variable from this group was associated with one of the two outcomes producing a HR statistically different from unity: previous Pap cytology. As can be seen in table 19, the hazard of high-risk HPV infection among women who reported having had 6 or more previous Pap smears is 0.46 (95%CI: 0.26-0.83) times the one for women who reported 0 to 5 previous Pap tests. However, this association became less protective and non-significant upon adjustment for age. Only one factor became significantly associated with high-risk HPV infection after controlling for the confounding effect of age: condom use. Condom use showed a protective effect for both high- and low-risk infections, but was statistically significant only for the former: age-adjusted HR for high-risk infection=0.55 (95%CI: 0.32-0.96) for women who ever used condoms compared to women who reported never having used them.

For incident low-risk HPV infection, a positive association was found with the use of non-commercial hygienic absorbents and with medical consultations for STDs. Frequent vaginal douching was found to be negatively associated with low-risk HPV infection. However, the confidence intervals for those HR did not exclude unity.

4.4 Crude and age-adjusted Cox models for selected lifestyle factors

As shown in table 20, smoking status, lifetime smoking exposure, and alcohol drinking resulted in positive associations with both outcomes in crude and age-adjusted models. However, none of these associations was statistically significant. Especially for smoking status, since it is very often correlated with sexual habits, further adjustment was necessary in order to determine whether there was an independent contribution to the risk of developing an infection.

Table 19: Crude and age-adjusted Cox proportional hazards regression models for various indicators of reproductive and contraceptive history

	Low-risk incident HPV infection				High-risk incident HPV infection			
	Crude		Age-adjusted		Crude		Age-adjusted	
Variables	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**
Years of OC use								
0.00-0.25 (1Q)	1.00	*p[1]=0.7974*	1.00	*p=0.9622*	1.00	*p=0.3573*	1.00	*p=0.8535*
0.26-3.00 (2Q)	1.03	(0.52-2.04)	1.09	(0.55-2.17)	0.56	(0.27-1.16)	0.52	(0.25-1.08)
3.01-7.00 (3Q)	1.18	(0.58-2.38)	1.16	(0.57-2.34)	0.98	(0.49-1.96)	1.19	(0.59-2.41)
7.01-29.00 (4Q)	1.05	(0.50-2.21)	1.00	(0.47-2.10)	0.52	(0.22-1.23)	0.83	(0.34-2.02)
0	1.00	*p=0.8500*	1.00	*p=0.7148*	1.00	*p=0.7458*	1.00	*p=0.4465*
< 6 years	1.11	(0.54-2.31)	1.16	(0.56-2.41)	0.68	(0.33-1.42)	0.68	(0.33-1.42)
≥ 6 years	0.98	(0.44-2.16)	0.93	(0.42-2.07)	0.80	(0.37-1.76)	1.27	(0.55-2.91)
Never	1.00		1.00		1.00		1.00	
Ever	1.06	(0.53-2.15)	1.07	(0.53-2.16)	0.73	(0.36-1.45)	0.81	(0.40-1.63)
Number of pregnancies								
0-1	1.00	*p=0.7820*	1.00	*p=0.8878*	1.00	*p=0.5530*	1.00	*p=0.2516*
2-3	0.93	(0.46-1.92)	0.84	(0.40-1.76)	1.07	(0.48-2.40)	1.63	(0.72-3.70)
4-5	0.76	(0.33-1.77)	0.66	(0.27-1.58)	0.82	(0.32-2.10)	1.50	(0.57-3.97)
≥ 6	1.20	(0.14-8.51)	0.98	(0.41-2.38)	0.86	(0.32-2.29)	2.09	(0.73-6.01)
0-2	1.00		1.00		1.00		1.00	
≥ 3	1.01	(0.61-1.67)	0.91	(0.53-1.54)	0.67	(0.39-1.16)	1.06	(0.59-1.92)

Table 19: **Crude and age-adjusted Cox proportional hazards regression models for various indicators of reproductive and contraceptive history (continued)**

	Low-risk incident HPV infection				High-risk incident HPV infection			
	Crude		Age-adjusted		Crude		Age-adjusted	
Variables	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**
Age at menarche								
8-11	1.00		1.00		1.00		1.00	
12-19	0.88	(0.49-1.57)	0.83	(0.46-1.49)	0.69	(0.38-1.25)	0.84	(0.46-1.53)
Use of hygienic tampons								
Never	1.00		1.00		1.00		1.00	
Ever	0.66	(0.27-1.65)	0.71	(0.28-1.77)	1.23	(0.55-2.72)	1.01	(0.45-2.25)
Use of non-commercial absorb.								
Never	1.00		1.00		1.00		1.00	
Ever	1.47	(0.90-2.40)	1.38	(0.83-2.28)	0.82	(0.47-1.43)	1.15	(0.65-2.04)
Vaginal douching								
Never/rarely	1.00		1.00		1.00		1.00	
Always/frequently	0.40	(0.13-1.28)	0.38	(0.12-1.23)	0.89	(0.35-2.24)	1.01	(0.40-2.55)
Vaginal/vulvar infection								
Never	1.00		1.00		1.00		1.00	
Ever	1.17	(0.61-2.23)	1.16	(0.61-2.23)	1.20	(0.58-2.45)	1.22	(0.60-2.51)
Genital discomfort past 5 yrs								
No	1.00		1.00		1.00		1.00	
Yes	0.98	(0.60-1.62)	0.97	(0.59-1.60)	0.67	(0.37-1.21)	0.71	(0.40-1.29)
Genital discomfort past 2 ds								
No	1.00		1.00		1.00		1.00	
Yes	0.74	(0.43-1.27)	0.73	(0.43-1.27)	0.90	(0.50-1.60)	0.90	(0.50-1.60)

Table 19: **Crude and age-adjusted Cox proportional hazards regression models for various indicators of reproductive and contraceptive history (continued)**

	Low-risk incident HPV infection				High-risk incident HPV infection			
	Crude		Age-adjusted		Crude		Age-adjusted	
Variables	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**
Condom use								
Never	1.00		1.00		1.00		1.00	
Rarely	1.01	(0.60-1.71)	1.03	(0.61-1.74)	0.58	(0.31-1.11)	0.52	(0.27-1.00)
Frequently	0.53	(0.25-1.13)	0.55	(0.26-1.17)	0.72	(0.36-1.44)	0.60	(0.30-1.21)
Never	1.00		1.00		1.00		1.00	
Ever	0.82	(0.50-1.34)	0.84	(0.52-1.38)	0.64	(0.37-1.09)	0.55	(0.32-0.96)
Consultation for STDs								
Never	1.00		1.00		1.00		1.00	
Ever	1.66	(0.96-2.89)	1.65	(0.95-2.86)	1.36	(0.71-2.59)	1.39	(0.73-2.66)
Previous Pap cytologies								
0-2 (1Q)	1.00	*p=0.5863*	1.00	*p=0.9683*	1.00	*p=0.0092*	1.00	*p=0.3661*
3-5 (2Q)	1.32	(0.65-2.71)	1.26	(0.62-2.59)	0.71	(0.37-1.35)	0.80	(0.42-1.54)
6-10 (3Q)	1.59	(0.77-3.27)	1.43	(0.69-3.00)	0.34	(0.14-0.81)	0.50	(0.20-1.21)
≥ 11 (4Q)	0.54	(0.54-2.56)	0.98	(0.44-2.22)	0.43	(0.19-0.94)	0.83	(0.34-2.03)
0-5	1.00		1.00		1.00		1.00	
≥ 6	1.17	(0.72-1.91)	1.06	(0.63-1.77)	0.46	(0.26-0.83)	0.71	(0.37-1.34)

1 P-value for trend for variables with 3 or more categories, obtained by treating the variable as ordinal in a Cox regression model.

Table 20: Crude and age-adjusted Cox proportional hazards regression models for selected lifestyle factors

	Low-risk incident HPV infection				High-risk incident HPV infection			
	Crude		Age-adjusted		Crude		Age-adjusted	
Variables	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**
Smoking status								
Not current smoker	1.00		1.00		1.00		1.00	
Current smoker	1.52	(0.93-2.48)	1.51	(0.92-2.48)	1.57	(0.91-2.72)	1.65	(0.95-2.86)
Lifetime smoking exposure (pack-years)								
0	1.00	*p=0.2391*	1.00	*p=0.3086*	1.00	*p=0.4425*	1.00	*p=0.1216*
0.01-1.40 (1Q)	0.85	(0.35-2.04)	0.90	(0.37-2.18)	1.26	(0.54-2.92)	1.01	(0.44-2.36)
1.41-5.40 (2Q)	1.65	(0.82-3.30)	1.71	(0.85-3.43)	1.00	(0.41-2.46)	0.97	(0.40-2.37)
5.41-11.20 (3Q)	1.65	(0.78-3.50)	1.62	(0.77-3.43)	1.39	(0.60-3.23)	1.77	(0.75-4.18)
11.21-55.55 (4Q)	1.20	(0.58-2.46)	1.13	(0.55-2.32)	1.29	(0.60-2.78)	1.79	(0.81-3.96)
0	1.00	*p=0.3459*	1.00	*p=0.4107*	1.00	*p=0.3879*	1.00	*p=0.1285*
0.01-4.50 (median)	1.33	(0.73-2.41)	1.41	(0.77-2.58)	1.13	(0.56-2.26)	0.97	(0.48-1.94)
4.51-55.55	1.29	(0.72-2.30)	1.23	(0.69-2.20)	1.32	(0.71-2.47)	1.76	(0.92-3.36)
Alcohol drinking								
Never	1.00		1.00		1.00		1.00	
Ever	1.42	(0.81-2.49)	1.45	(0.82-2.56)	1.28	(0.69-2.36)	1.26	(0.68-2.32)

1 P-value for trend for variables with 3 or more categories, obtained by treating the variable as ordinal in a Cox regression model.

4.5 Multivariable Cox regression models for low-risk HPV infection

Variables that were considered as potential determinants for low-risk HPV infection based on crude and age-adjusted analyses were as follows: age, race, education level, age at first sexual intercourse, number of sexual partners (lifetime, past 5 years, and past year), oral sex, use of non-commercial sanitary absorbent during menses, vaginal douching, genital discomfort in the past 2 days (pain, itching, discharge), medical consultation for STDs, and smoking status. Women with complete information for these variables were included in the analysis for the selection of the final model (N=628). A total of 65 incident low-risk HPV infections was detected during follow-up.

Mutual adjustment of the determinants identified in crude and age-adjusted models had a substantial effect in reducing the magnitude of association and statistical significance for variables that had shown borderline significance or that were not statistically significant. For example, once further adjusted for other markers of sexual activity, the associations for lifetime number of partners and cunnilingus were moved towards unity and clearly lost their statistical significance. An increase in the magnitude of association for the use of non-commercial absorbents during menses was noted upon adjustment for level of education and for number of sexual partners of the past 5 years.

Table 21 shows the final models for incident low-risk HPV infection. The determinants identified with the unrestricted sample, i.e. including infections with unknown HPV types as low-risk HPV infections, were age, number of sexual partners in the past 5 years, and level of education attained. In terms of age, women in any of the last 3 age categories had a hazard of low-risk infection during follow-up 1.41 (95%CI: 1.06-1.86) times higher than women from the preceding age category after controlling for number of sexual partners in the past 5 years and education level.

Sexual partners during the past 5 years, rather than lifetime partners was a strong determinant of low-risk HPV infection. Women who had 2 or more sexual partners in the past 5 years had

a risk of developing a low-risk infection 3.51 times higher than women who had 0 or 1 partner during this period (95%CI: 2.12-5.82) after controlling for age and education level. Higher education level was associated with an elevated hazard for a low-risk HPV infection: adjusted HR=1.73 (0.98-3.07) for women with at least some high school education compared to women with at most a completed elementary level schooling.

Low-risk HPV infections were further defined as only those infections with known types. Any HPV infection with an undefined type was therefore excluded from the analysis resulting in a sample size of N=613 and 49 incident infections. The same variables previously identified were again selected as relevant in this final model, with the addition of a new variable: use of non-commercial sanitary absorbent during menses. Women who reported having ever used non-commercial sanitary absorbents had a higher hazard of developing a low-risk HPV infection during follow-up: 1.82 (95%CI: 1.01-2.28). The association with age remained unchanged. Both the associations with number of sexual partners and education became stronger in magnitude.

Table 21: Final multivariable Cox regression models for incident low-risk HPV infection

	Including HPV of unknown types		Excluding HPV of unknown types	
Variables	**Hazard ratio**	**95% CI**	**Hazard ratio**	**95% CI**
Age[1]	1.41	(1.06-1.86)	1.40	(1.01-1.95)
Number sexual partners past 5 years				
0-1	1.00		1.00	
≥ 2	3.51	(2.12-5.82)	4.49	(2.52-8.00)
Education level				
≤ elementary	1.00		1.00	
≥ high school	1.73	(0.98-3.07)	2.22	(1.15-4.27)
Use of non-commercial hygienic absorbents				
Never			1.00	
Ever			1.82	(1.01-3.28)

1 Age coded into 4 categories (18-24, 25-34, 35-44, 45-60) and treated as an ordinal variable in the model

4.6 Multivariable Cox regression models for high-risk HPV infection

Variables considered as potential explanatory factors for the multivariable Cox regression model for high-risk HPV infection were identified using the results from the crude and age-adjusted analyses. Any variable with a p-value less than 0.25 or biological relevance was considered. The potential explanatory variables are as follows: age, age at first sexual intercourse, number of sexual partners (lifetime, past 5 years, past year), anal intercourse, oral sex, age at menarche, number of pregnancies, condom use, vaginal discomfort in the past 5 years (pain, itching, discharge), previous Pap tests, and smoking status. The sample used for these analyses included 624 women with complete information for the above-mentioned variables and gave rise to 52 incident high-risk HPV infections.

In a similar fashion to what has been seen for markers of sexual activity when building the models for low-risk HPV infection, recent number of sexual partners (past year) was identified as an independent determinant of risk rather than lifetime partners. Although an association of high-risk infection with anal intercourse and cunnilingus was suggested by results from crude and age-adjusted analyses, it completely disappeared upon adjustment for age at first intercourse and number of sexual partners. The age-adjusted association seen with current smoking was also decreased after controlling for the effect of number sexual partners, corroborating the hypothesis that smoking status is correlated with sexual behaviour. Smoking was therefore not included in the final model.

As shown in table 22, the variables identified as independent determinants of incident high-risk HPV infection were age, age at first intercourse, number of sexual partners in the past year, and condom use. In terms of age, women in any of the last 3 age categories were less likely than women from the preceding age category to have developed a high-risk HPV infection during follow-up, after controlling for age at first sexual intercourse, number of sexual partners in the past year, and condom use: HR=0.55 (95%CI: 0.38-0.79) for women in one age category compared to women in the preceding one. An older age at first sexual intercourse was also found to be protective. Women who had their first sexual intercourse at

18 years or older had a hazard of developing a high-risk infection of one-third compared to that of women who initiated sex at 17 years or younger (HR=0.36, 95%CI: 0.18-0.72). Women with 2 or more sexual partners in the past year had a 4.47 times higher risk of incident infection with a high-risk type during follow-up than women with 0 or 1 sexual partner, after controlling for other variables included in the final model (95%CI: 2.16-9.26). A protective association was obtained for condom use: $HR_{adj.}$=0.57 (95%CI: 0.33-0.98) for women who reported having ever used condoms compared to women who had never used them.

Table 22: Final multivariable Cox regression model for incident high-risk HPV infection

Variables	Hazard ratio	95% CI
Age[1]	0.55	(0.38-0.79)
Age at first sexual intercourse		
≤ 17 years	1.00	
≥ 18 years	0.36	(0.18-0.72)
Number sexual partners past year		
0-1	1.00	
≥ 2	4.47	(2.16-9.26)
Condom use		
Never	1.00	
Ever	0.57	(0.33-0.98)

1 Age coded into 4 categories (18-24, 25-34, 35-44, 45-60) and treated as an ordinal variable in the model

1 Advantages of studying incident infections

Most studies of HPV infection have been based on cross-sectional designs, mainly because of the cost associated with prospective cohort studies. Although limited by the absence of temporal relationship, these prevalence studies have contributed enormously to the knowledge of the association between HPV infection and cervical cancer. However, in view of primary prevention of cervical neoplasia, understanding the incidence of clinically relevant HPV would allow to target the early causes of disease (Franco, 1997). More specifically, understanding the transmission of low-risk and high-risk HPV is a first step towards the development of public health programs aimed at preventing HPV infection.

2 Population epidemiological profile

The epidemiological profile obtained was one that could be expected for a non-industrialized, Latin American population. The low level of education – only 18% of the study women having attended high school – and the low income – the median monthly family income being US$265 – are among the expected sociodemographic features. Another characteristic is the relatively low number of sexual partners reported, the median being 2 lifetime sexual partners. Similar figures have been obtained for lifetime sexual partners in a previous study in Joa Pessoa, Brazil (Franco et al., 1995). As expected in a non-industrialized country where birth rates are historically higher than in industrialized countries, the number of pregnancies reported was high: close to two-thirds of the women reported having had 3 or more pregnancies. The relatively low use of hygienic tampons and condoms, combined with the common use of non-commercial absorbents during menses brought greater strength to the picture of an economically disadvantaged population.

The age distribution obtained ranged from 18 to 59 years, with a median of 33 years. The majority of women were aged between 25 and 44 (70%) and 17% were younger. Such an age distribution should enable us to study the effect of age on the incidence of HPV infection.

3 Incidence density rates of HPV infection

To our knowledge there exists only one other study in which person-time incidence rates of HPV infection have been published (van Doornum et al.,1994). The authors of this study conducted in Amsterdam reported an incidence rate of 47.1 HPV infections per 100 women-years, or 39.2 infections per 1000 women-months. This incidence rate is therefore more than twice as high as the one obtained in the current study for all HPV infections irrespective of type: 16.4 newly detected infections per 1000 women-months (table 11). However, many differences between the two studies render the comparison difficult to interpret. The populations themselves are totally different: the Brazilian women represent a population from a developed country with few sexual partners, whereas the women who participated to the Amsterdam study had had at least 5 sexual partners in the 6 months prior to enrollment. This inclusion criterion was related to the primary object of the study: the heterosexual spread of HIV (van Doornum et al.,1994). The Ludwig-McGill study tested for more than 40 HPV types whereas the Amsterdam study was designed to detect only types 6/11, 16, 18, and 33, a difference that would intuitively result in higher incidence rates of infection in the Brazilian cohort. In the Amsterdam study, however, HPV DNA was detected from cell specimens taken at multiple body sites such as mouth, anus, rectum, cervix, labia minora (van Doornum et al.,1994). A fairer comparison would require a breakdown per site for the Amsterdam study and restriction to types 6/11, 16, 18, and 33 in the Brazilian study. Even these modifications would leave room for a lot of variation between the 2 populations.

The presence of HPV DNA was assessed every 4 months. It is possible that some women have developed and cleared HPV infections during these intervals. The incidence density rates obtained would then underestimate the true incidence density in the study population. Another possible source of underestimation would be the assumption that had to be made with regards to the timing of infection. For example, when a woman became HPV positive at t_2, the positivity was considered at that exact point in time, not in the middle of the interval t_1 - t_2. However, both of these sources of underestimation of the incidence density are minimized by

the relatively frequent testing: 4 months compared to 6 months or even 12 months in other studies.

By definition, women who harboured an HPV infection at the baseline visit were excluded from this research project. Some of these women might be more susceptible to HPV infection and be at higher risk of developing new infections than women who were HPV-free at baseline. This would also contribute to underestimate the incidence density rates. However, since the dynamics of infection have barely been studied in detail, it is difficult to estimate the risk of subsequent infection conveyed by an HPV positive status at one point in time. This question could be addressed by following closely the HPV status of these women for clearance of the infection, development of new infections, and their respective timing. As the ongoing Ludwig-McGill cohort study progresses, it will be possible to study the dynamics of infections (development and clearance) with a larger number of subjects and a longer follow-up time. This will allow to include in a future study of incidence, women who were HPV positive at baseline and who became HPV negative at a subsequent visit. Such an analysis could allow to determine if these women were, in fact, at higher risk of developing subsequent HPV infections.

Different patterns for low-risk and high-risk HPV infection are suggested by the incidence rates obtained after stratification for age. The incidence rates for low-risk HPV infection increase slightly with age, although not significantly. Conversely, incidence rates for high-risk HPV infections decreased markedly with age. The reasons for these differences can only be hypothesized at this point. The immune response to HPV infection could differ between low-risk and high-risk types. It has been hypothesized that antibodies might be less efficiently produced as a result of transient infections (Galloway, 1996). Given results indicating that high-risk HPV types were more likely to be persistent than low-risk ones (Hildesheim et al., 1994), it could be further hypothesized that the antibody production might be more efficient for high-risk HPV. In this model, older age would therefore result in a better immune response due to a longer lifetime exposure to HPV. These patterns could also be indicative

of separate pathways of transmission for low-risk and high-risk types. If it is considered that younger women have a higher exposure in terms of new sexual partners, the incidence rates obtained could suggest that transmission of high-risk types is more strongly associated with sexual activity than that of low-risk types, given the decrease in incidence rates with age for high-risk but not for low-risk types.

The effect of HPV DNA detection frequency on the incidence rates of infection was an important feature to verify, since the obvious criticism of studies with multiple HPV testing is that they would falsely increase the incidence rates of infection by allowing to detect a number of events of interest that would not have been detected otherwise. However, this was not apparent in our study. Upon stratification for the number of returns after the baseline visit (1, 2 or 3), all the stratum-specific incidence rates of overall HPV infection were comparable (table 12). This suggests that while the cumulative number of infection is elevated by multiple HPV testing during the course of the study, the incidence density rates are not affected. Comparisons of incidence rates among studies would consequently not be hampered by differences in their HPV testing schedules. In terms of study design, the implications of this finding are substantial. Prospective studies on HPV would not be constrained to use the same testing intervals in order to obtain incidence rates that can be fairly compared with those from other studies.

4 Cumulative distribution of low-risk and high-risk HPV infection

The actuarial analysis of newly detected HPV infections resulted in similar cumulative probabilities for both low-risk and high-risk types (figure 3). No comparable analysis has been published in the literature and therefore these results cannot be contrasted with others. However, since they were obtained for only a sub-group of the Ludwig-McGill cohort study population, it will be interesting to compare the present estimates with those obtained in the future for all participants and with a longer follow-up.

5 Determinants of incident HPV infection with low-risk types

5.1 Sociodemographic characteristics based on crude and age-adjusted analyses

Results from Kaplan-Meier analyses for sociodemographic characteristics suggested that only ethnicity was associated with time to low-risk HPV infection (table 13). Literature on HPV suggests that the susceptibility to HPV infection does not vary according to race (Schiffman, 1994). The higher likelihood of incident low-risk HPV infection among non-white women could be associated with factors such as diet, socioeconomic status or behaviour for which ethnicity would only be a proxy.

A greater hazard of low-risk HPV infection was also found among women who reported a higher level of education (table 17). This finding is opposite to those published from prevalence studies. In two American populations, women who had attained a higher level of schooling had a lower prevalence of HPV infection than the less educated women (Hildesheim et al., 1993; Bauer et al., 1993).

The direction of association obtained with age was also different from what was expected based on prevalence studies of overall HPV (Hildesheim et al., 1993; Bauer et al., 1993) and a cross-sectional study where low-risk infections were treated separately from high-risk ones (Richardson, 1996). Instead of a decrease of incidence with increasing age, older women in the current study had a slightly higher hazard of infection with a low-risk HPV type.

5.2 Markers of sexual activity based on crude and age-adjusted analyses

Based on actuarial analysis, women with multiple partners in the past 5 years, in the past year, and women who had experienced cunnilingus had higher cumulative rates of newly detected HPV infection with low-risk types (table 14). In regression models, adjustment for age increased the magnitude of association with cunnilingus, also rendering it statistically significant (table 18).

The association between number of sexual partners and prevalence of HPV infection has varied greatly among studies conducted with different populations. Results from these studies have ranged from an absence of association to a strong relationship between multiple sexual partners and prevalence of HPV infection (Ley et al.,1991; Bauer et al.,1993; Wheeler et al.,1993; Rohan et al.,1991; Hildesheim et al.,1993; Kjaer et al.,1993). However, recent results from our group suggested that the prevalence of low-risk HPV types was not associated with sexual activity (Franco et al.,1995; Richardson,1996). The higher incidence of low-risk HPV infection among women with multiple sexual partners, which was obtained in the present study, is in contradiction with these previous results from our group.

No association was obtained in this study between the occurrence of low-risk HPV infection and the age at first intercourse, one of the frequently used markers of sexual activity. Similar results have been generated for low-risk HPV types in cross-sectional studies conducted by our research team in Brazil (Franco et al.,1995) and in Montreal (Richardson,1996).

5.3 Indicators of reproductive and contraceptive history based on crude and age-adjusted analyses

Results from the actuarial analysis suggested that three variables from this group were associated with the cumulative distribution of newly detected infections with low-risk HPV types (table 15). Use of non-commercial absorbents during menses and consultation for STDs were both associated with higher cumulative probabilities of infection, whereas the reverse was seen with frequent vaginal douche. Similar results were obtained by Cox regression analysis, with very little variation upon adjustment for age (table 19). Few studies have addressed the use of non-commercial absorbents during menses, a factor that is obviously more appropriate for populations from non-industrialized countries. Previous results from our research team had not suggested an increased prevalence of low-risk HPV infection associated with the use of home-made absorbents (Franco et al., 1995). However, our results for vaginal douche corroborate those from two studies that suggested a decreased prevalence of low-risk HPV infection among women who used vaginal douches (Franco et al., 1995; Richardson, 1996).

5.4 Lifestyle factors based on crude and age-adjusted analyses

Results from both Kaplan-Meier and Cox regression analyses suggested an elevated hazard of infections with low-risk HPV types for current smokers and for women who reported drinking alcohol (tables 16 & 20). The lifetime smoking exposure did not have an effect on the incidence of low-risk HPV infection. The elevated hazard of low-risk HPV infection for current smokers is in accordance with the hypothesis of tobacco-induced immunosuppression, which could result in a less efficient ability to protect the cervix against infection by HPV (Barton et al., 1988). Due to the transient nature of HPV infections, an effect of current rather than lifetime exposure to tobacco is intuitively plausible.

5.5 Determinant profile of incident low-risk HPV infection based on multivariable Cox regression analysis

The final model obtained included age, number of sexual partners in the past 5 years, and level of schooling attained as independent predictors of incident low-risk HPV infection (table 21).

As previously described, the positive association between age and time to development of a low-risk infection was unexpected. Possible explanations for this relationship with age include a lesser efficiency of the immune response to low-risk than to high-risk HPV types, and a lesser importance of sexually-related transmission for low-risk types compared to high-risk HPV types. Different patterns of latency and reactivation of infection according to oncogenicity could also be considered.

The strong independent association between the number of sexual partners and the risk of occurrence of low-risk infection suggests a transmission via sexual routes. It is in contradiction with recent results which led to the hypothesis of a lesser importance of sexual transmission for low-risk types (Franco et al.,1995; Richardson,1996).

The positive association between the level of schooling and the time to low-risk HPV infection was unexpected in view of results from prevalence studies. Published results suggest a lower, although not statistically significant, prevalence of overall HPV infection among women with a higher education level (Hildesheim et al.,1993; Bauer et al.,1993). The two American populations from these studies differ markedly from the Brazilian population studied in the present research project. Results from case-control studies of invasive cervical cancer also showed a decrease in HPV prevalence within increasing levels of schooling among control women (de Sanjose et al.,1996). The level of education is used as a marker of socioeconomic status and is often associated to behavioural differences related to various topics which include: sexual habits, diet, attitude towards health and screening. The effect of education can possibly vary depending on the countries' social and economical situation. This makes education a very global marker and adds to the complexity of its interpretation.

The multivariable analysis was also conducted on a restricted sample from which all women with HPV of unknown type were excluded. This procedure was performed in order to eliminate possible misclassification due to the inclusion of non-typed HPV infection into the low-risk category. The analysis allowed a refinement of the determinant profile for low-risk HPV types. In addition to the three variables mentioned above, use of non-commercial absorbents during menses was included in the final model. While the magnitude of association between age and time to low-risk infection did not change, the hazard ratios related to the number of sexual partners in the past 5 years and to the level of education increased, as well as their respective statistical significance.

The profile obtained with this restricted sample suggests that although sexual activity is an independent determinant, behavioural and hygiene components could possibly have an impact on the likelihood of developing a low-risk HPV infection. These results might indicate a more complex and multi-factorial mode of transmission for low-risk HPV types than for high-risk ones, as will be described in the next section.

6 Determinants of incident HPV infection with high-risk types

6.1 Sociodemographic characteristics based on crude and age-adjusted analyses

The only sociodemographic characteristic associated with the distribution of high-risk HPV infection during follow-up was age (table 13). Results from Cox regression models confirmed a strong negative association between age and time to high-risk HPV positivity (table 17). The direction of association is consistent with results from studies which had suggested that the prevalence of HPV infection decreased markedly with age (reviewed in Schiffman, 1994). The reasons for this important drop of both prevalence and incidence of infection with high-risk HPVs are not known. It is possible that the immune response to high-risk HPV is more efficient in women who have had a longer exposure to them. Given the opposite relationship obtained with age for incident infections with low-risk HPV types, such a hypothesis implies

a different reaction of the immune system towards low-risk HPVs. A less efficient antibody production in response to transient infections, a characteristic more often associated with low-risk HPV types, has already been hypothesized (Galloway, 1996).

6.2 Markers of sexual activity based on crude and age-adjusted analyses

All the markers of sexual activity were associated with the distribution of incident high-risk HPV infection over time (table 14). Results from the age-adjusted Cox regression models suggest a stronger association of incident high-risk infection with the past year sexual activity rather than with the past 5 years or lifetime number of sexual partners (table 18). Considering that the latency period is much shorter for the development of HPV infection than of neoplastic lesions, it is conceivable that the most recent sexual history could be a better predictor of the risk of HPV infection. In a cross-sectional study using new sexual partners for 3 reference periods (past year, past 1-5 years, past 5-10 years), recent sexual activity was a better predictor of HPV prevalence than distant sexual history (Fairley et al., 1994).

6.3 Indicators of reproductive and contraceptive history based on crude and age-adjusted analyses

Two indicators of reproductive and contraceptive history were associated with the distribution of newly detected HPV infections with high-risk types in crude analyses (tables 15 & 19). Condom use and a higher number of previous Pap tests were both found to be protective.

Although considered as a very efficient barrier for most sexually-transmitted diseases when used correctly and during the entire sexual relation, the role of condoms towards HPV infection has yet to be determined. One study found lifetime condom use as an independent protective determinant of prevalence of high-risk HPV infections. Conversely, recent condom was independently associated with an increased prevalence of low-risk HPV infections (Richardson, 1996).

The protective association for higher number of previous Pap smears might be interpreted as resulting from a group of women who are generally more health-conscious than women who reported a lesser use of screening procedures. However, this apparent association was confounded by age, this variable being correlated with the number of Pap smears and independently associated with the outcome. Results from the age-adjusted model for previous Pap smears, which show a milder protective effect of previous cytologies and a loss of statistical significance compared to the crude model, are consistent with this idea.

6.4 Lifestyle factors based on crude and age-adjusted analyses

Results generated by the crude and age-adjusted analyses suggested slightly elevated hazards of high-risk HPV infection for current smokers compared to non-current smokers (tables 16 & 20). Some studies of prevalent HPV infection have suggested an increased risk associated with smoking, based on crude analyses. Further adjustments for sexual activity and sociodemographic indicators either lead to loss in magnitude and statistical significance (Bauer et al.,1993; Hildesheim et al.,1993; Ley et al.,1991), to reversal of the direction of association and loss of significance (Wheeler et al.,1993) or to identification of smoking as an independent determinant of prevalence of HPV infection (Rohan et al.,1991).

6.5 Determinant profile of incident high-risk HPV infection based on multivariable Cox regression analysis

The independent determinants of time to high-risk HPV infection as determined by multivariable Cox modelling were: age, age at first sexual intercourse, number of sexual partners in the past year, and condom use (table 22). The protective association with age remained, although weaker, after controlling for markers of sexual activity in the model. These results indicate that the drop in hazard of high-risk infection with increasing age can be partly, but not entirely, attributed to lower sexual activity among older women.

Similarly to what has been seen for markers of sexual activity when building the models for low-risk HPV infection, recent number of sexual partners (past year) was found as an independent determinant of risk rather than lifetime partners.

Condom use was found to be protective against acquisition of high-risk HPV infection. It remains to be determined whether the identification of condom use as an independent protective factor for high-risk but not for low-risk infection is indicative of differences in their modes of transmission.

The determinant profile obtained for high-risk HPV infection is similar to the one for low-risk, since it includes both age and number of sexual partners. However, older age is associated with a higher hazard for low-risk infections and a lower one for high-risk ones. In addition, the variables quantifying the sexual partners do not refer to the same period of reference. The determinant profile for incident infection with high-risk HPV types is strongly related to sexual activity. In addition to being associated with sexual activity, incidence of low-risk HPV infection was related to socioeconomic status and to the use of non-commercial absorbents during menses. This could be seen as partly supporting the hypothesis of different modes of transmission for low-risk and high-risk HPV (Franco et al.,1995; Richardson,1996). However, since sexual activity is evidently an important determinant of incidence of low-risk HPV infection, no clear conclusion can be drawn on this topic.

7 Limitations and strengths of the study

7.1 Interview-derived information

The sexual behaviour of the male partner has not been taken into account in this study. This marker of sexual behaviour is the most difficult to obtain, since it requires full participation from the husbands or male partners of the participating women. Case-control studies that compared the sexual behaviour of male partners have been plagued by low response rates. Nevertheless, they have indicated that husbands of women with cervical cancer were more likely to report multiple extramarital partners and contacts with prostitutes than husbands of control women (Brinton et al.,1989; Kjaer et al.,1991; Bosch et al.,1996). In order to obtain information on male partners' sexual behaviour, it would have been necessary to recruit the husbands of the women enrolled in the Ludwig-McGill study. Although it would have undoubtedly provided insightful information, this would have resulted in a substantial increase in necessary funding, for an already expensive long-term follow-up study. The study was therefore designed to focus on women only.

The baseline interview elicits information on lifetime use of oral contraceptives, rather than current or recent use. In studying the incidence of neoplastic lesions, long-term oral contraceptive use is probably an appropriate factor to investigate due to the latency normally involved in such a pathological effect. For incident HPV infection, however, the most useful information might be current use or fairly recent use of oral contraceptives. The Ludwig-McGill cohort study will address this question, since information on oral contraceptive use since the last visit is collected at each follow-up return. As data from more women become available for follow-up visits, recent oral contraceptive use will be studied in relation to HPV infection.

The information collected on condom use could have been more detailed. The relative frequency of use is only one aspect of this complex behaviour. More qualitative information would be needed to perform a thorough assessment of condom use as a determinant of HPV

infection. This information could encompass the moment at which the condom is used (during the entire intercourse or not), and consistency of use (Richardson, 1996). The population studied is not the best one for addressing the various dimensions of condom use, since few women reported using them regularly. Younger women with higher levels of education might prove to be more useful in learning about the effect of condom on HPV infection.

There are advantages and disadvantages in performing an interview instead of a written questionnaire. Some believe that sensitive information, e.g. sexual behaviour, would be better obtained if the women were writing it instead of telling an interviewer. The fact that the study was conducted in a non-industrialized country with a low level of schooling prompted the investigators to use interviews performed by female nurses. The nurse interviewer can therefore ask the questions in simple terms and make sure it is well understood. The use of an interview rather than a written questionnaire might also incite women to leave a smaller number of unanswered questions. Furthermore, it ensures that a reasonable amount of time is spent providing answers, which is not necessarily the case when a questionnaire is handed out to the women.

7.2 Limitations of the assessment of negative HPV status at baseline

For this research project, women who had HPV negative cervical cell samples at the baseline visit were considered free of HPV infection and therefore at risk for incident ones. It has been proposed that insufficient knowledge on latency and reactivation of HPV could result in falsely classifying women as negative if the assessment was based on only one HPV test (Schiffman, 1994). The solution proposed was to study the incidence of HPV infection in young women who have not been sexually active. The recruitment of a young population which would require parental consent raises ethical and logistical issues especially when working on viruses transmitted mainly via sexual contacts. Furthermore, the identification of cervical infections or lesions in such a population could easily become problematic.

The current study of incident HPV infection based on negative HPV status at the baseline visit was our first effort in trying to elucidate the dynamic role of HPV. This definition of HPV "negativity" may have different implications for low-risk and high-risk HPV types, due to possible differences in their likelihood of transience and persistence (Hildesheim et al., 1994; Franco et al.). This definition might therefore be more appropriate for high-risk HPV types, since they are more likely to be persistently detected than low-risk types.

7.3 Misclassification of exposure and outcome

Some misclassification might be present in the exposure variables. However, since women are interviewed without knowing their HPV status, it is more likely that misclassification would be non-differential, resulting in a bias towards the null for the measure of association.

Misclassification of HPV status due to imperfect sensitivity and specificity is expected since no diagnostic test is perfect. However, PCR-based techniques have higher sensitivity and specificity for detection of HPV DNA than hybridization techniques used before the 1990s which did not amplify the viral DNA (Franco, 1997). They have been shown to result in less misclassification of HPV status than previous methods used in epidemiologic studies of HPV infection and cervical cancer (Schiffman and Schatzkin, 1994). Furthermore, any misclassification due to the detection method would be non-differential.

The inclusion of unknown HPV types into low-risk HPV would likely result in misclassification of the low-risk HPV outcome variable. The analysis of low-risk incident HPV infections excluding unknown types showed a clearer determinant profile, suggesting that the associations had in fact been diluted by this misclassification.

7.4 Internal and external validity

Given the important involvement required from the participants, it was not possible to conduct this study with a sample of individuals from the general population. To preserve the

internal validity of the study, completeness of information and a good follow-up return rate had to be favoured over the representativeness of the study sample towards the general population. The subjects were therefore recruited as a convenience sample of volunteer women who presented themselves at the participating clinics during the accrual period. The participation rate, 70.5% and the return rate of 80% after 36 months are high considering the required frequency of visits, the physical examinations, and the lengthy questionnaires. As more data become available, it will be possible to assess the possibility of a selection bias using basic sociodemographic information from women who refused to participate and some of the baseline information from women who did and did not comply with the scheduled returns.

In terms of external validity, the incidence density rates cannot be generalized at the population level. Until more is known about the dynamics of HPV infection and the important risk factors, these rates will remain specific to the study population, but can serve as comparison figures when similar rates become available from other studies. In terms of risk factors for incident HPV infection, however, there is no evidence to this date indicating a different etiology or transmission of infection in certain populations. These results are thus more readily generalizable, with the caution that they were found in a population with a relatively low level of education, low socioeconomic status, and in a non-industrialized country.

7.5 Strengths of the study

This prospective cohort study is the first one to characterize the dynamic process of occurrence and clearance of HPV, based on multiple HPV testing at predefined time intervals. The current research project is a first glance at the possible determinants of incident HPV infections. It is based both on a subset of the entire study population and on a limited view of the follow-up information that is due to accumulate in the study. More insightful analyses will be possible as more HPV results are available and as follow-up accumulates. Interview information will also provide current exposure for various potential determinants.

8 Future directions

Larger epidemiological studies focussing on cervical HPV infection now allow epidemiologists to study separately low-risk and high-risk HPV types. HPV infection results in different pathological manifestations according to oncogenicity. Recent prevalence studies suggested that not only their pathological manifestations might be different, but also their modes of transmission (reviewed in Franco,1997). High-risk HPVs would tend to be more strongly associated with markers of sexual activity than low-risk HPV types.

In order to address such a question with regards to incident HPV infection, an analysis strategy, different than the one used for the current research project, might be used. This new strategy could possibly tackle the imperfect choice of HPV negative status at baseline as an indication of the absence of HPV infection. Based on data from the Ludwig-McGill prospective cohort study, this new strategy could consist in a case-control analysis that would be done separately for low-risk and for high-risk HPV infection. The control subjects could be women with negative HPV DNA detection tests for a defined number of follow-up visits. Case subjects could be women with at least one instance of HPV positivity for the same defined follow-up. By reducing misclassification of the HPV status at baseline, this analysis strategy might help unveil possible differences in the modes of transmission of genital HPV infection.

CONCLUSION

In summary, this research project has permitted to:

1) Calculate the incidence density rates for overall, low-risk, and high-risk cervical HPV infections, which were respectively: 16.4, 9.3, and 7.6 per 1000 women-months. Different patterns were seen upon stratification for age: a slight increase of the incidence rates of low-risk HPV infection with age, and an evident decrease of the incidence rates of high-risk HPV infection with age.

2) Identify independent determinants of incident low-risk genital HPV infection. These determinants were age, number of sexual partners in the past 5 years, education level, and use of non-commercial hygienic absorbents during menses.

3) Identify independent determinants of incident high-risk genital HPV infection. These determinants were age, age at first sexual intercourse, number of sexual partners in the past year, and condom use.

In conclusion, results from this research project do not corroborate the absence of association between low-risk HPV types and markers of sexual activity, as found in recent studies. They rather suggest differences in correlates of incident infection with low-risk and high-risk HPV types: a multi-factorial determinant profile for non-oncogenic HPVs as compared to a strong sexually-transmitted profile for oncogenic types.

Further elucidation of the modes and dynamics of transmission of genital HPV infection, specifically for oncogenic types, would contribute to primary prevention in both industrialized and developing countries by allowing the orientation of public health programs towards efficient education and intervention efforts to ultimately reduce risk of cervical cancer (Franco, 1997).

ACKNOWLEDGMENTS

I wish to thank Dr. Eduardo Franco without whom none of this would have been possible. His passion for cancer epidemiology, made him a very enthusiastic supervisor. His trust and unconditional support allowed these two years to be a wonderful learning experience. I am extremely grateful for having had the opportunity to work under his supervision.

I would also like to thank Dr. Luisa Villa and her team at the Ludwig Institute for Cancer Research in Brazil, for their dedicated work "behind the scenes". The HPV data that was so easily made available to me represented countless hours of laboratory work.

Dr. Abdissa Negassa provided advice and feedback concerning the statistical analysis portion of this work. I am very grateful for his patience and availability.

Dr. Javier Pintos was generous enough to offer me a private introduction to the "SPSS World". After guiding my first steps, he was always available to answer my technical and analytical questions. He also provided most helpful feedback on this document and lots of encouragement. Harriet Richardson very kindly took some time to read my thesis and provided invaluable scientific and grammatical suggestions. She was always there for fruitful discussions and moral support. Sincere thanks to two wonderful colleagues and dear friends.

I am indebted to Rita Zakarian, who was responsible for the database management in Montreal. She assisted me in various ways, including the translation of the original questionnaire from Portuguese to English, sometimes even to French. Candida Pizzolongo, our Division's secretary, relieved me of administrative work on many occasions and was always eager to help when I needed her. My most sincere thanks to both of them.

It has been a pleasure to share the "thesis experience" with Sophie Kulaga, my office neighbour and colleague. Her collaborative attitude made me feel part of a team, even during

the inevitable isolated moments spent analysing and writing. She also shared with me any useful piece of information related to the thesis writing and submission, which often made my life much easier. I admired her motivation and dedication. It was a great privilege to work with her.

Finally, my very special thanks to Benoît, my husband. He revived my self-confidence in times of doubt and showed strong support from beginning to end. His love for science and his genuine interest in my work were constant sources of inspiration.

I am also grateful to my family and my parents in-law for their support and interest in my endeavours. Their words of encouragement gave me the strength I needed to reach the finish line.

FINANCIAL SUPPORT

My involvement in this study was made possible in part by Health Canada through a National Health Research and Development Program (NHRDP) Research Training Award. Additional financial support was awarded to me by the Fondation Desjardins, programme Girardin-Vaillancourt.

The Ludwig-McGill cohort study is funded by the Ludwig Institute for Cancer Research, the Medical Research Council (MA-13647) and the National Institutes of Health (R01 CA70269-01).

BIBLIOGRAPHY

Armitage, Berry. *Statistical Methods in Medical Research*. Third Edition ed. Oxford: Blackwell; 1994.

Baken, L.A., Koutsky, L.A., Kuypers, J., Korosok, M.F., Lee, S.K., Kiviat, N.B., and Holmes, K.K. Genital human papillomavirus infections among male and female sex partners: prevalence and type-specific concordance. *J Infect Dis*. 1995;171:429-432.

Baker, C.C. and Howley, P.M. Differential promoter utilization by the bovine papillomavirus in transformed cells and productively infected wart tissues. *EMBO Journal*. 1987;6:1027-35.

Banks, L. and Matlashewski, G. Biochemical and biological activities of the HPV E5 proteins. In: Lacey, C., (ed.). *Papillomavirus Reviews: Current Research on Papillomaviruses*. London: Leeds University Press; 1996:39-45.

Barton, S.E., Jenkins, D., Cuzick, J., Maddox, P.H., Edwards, R., and Singer, A. Effect of cigarette smoking on cervical epithelial immunity: a mechanism for neoplastic change? *Lancet*. 1988;2:652-654.

Bauer, H.M., Hildesheim, A., Schiffman, M.H., Glass, A.G., Rush, B.B., Scott, D.R., Cadell, D.M., Kurman, R.J., and Manos, M.M. Determinants of genital human papillomavirus infection in low-risk women in Portland, Oregon. *Sex Transm Dis*. 1993;20:274-278.

Bauer, H.M., Ting, Y., Greer, C.E., Chambers, J.C., Tashiro, C.J., Chimera, J., Reingold, A., and Manos, M.M. Genital human papillomavirus infection in female university students as determined by a PCR-based method. *JAMA*. 1991;265:472-477.

Beral, V., Hannaford, P., and Kay, C. Oral contraceptive use and malignancies of the genital tract. *Lancet*. 1988;2:1331-1335.

Berkhout, R.J.M., Tieben, L.M., Smits, H.L., Bouwes Bavinck, J.N., Vermeer, B.J., and Schegget, J.T. Nested PCR approach for detection and typing of epidermodysplasia verruciformis-associated human papillomavirus types in cutaneous cancers from renal transplant recipients. *J Clin Microbiol*. 1995;33:690-695.

Bernard, B.A., Bailly, C., Lenoir, M.C., Darmon, M., Thierry, F., and Yaniv, M. The human papillomavirus type 18 (HPV18) E2 gene product is a repressor of the HPV18 regulatory region in human keratinocytes. *J Virol*. 1989;63:4317-4324.

Bernard, H.U., Chan, S.Y., Manos, M.M., Ong, C.K., Villa, L.L., Delius, H., Peyton, C.L., Bauer, H.M., and Wheeler, C.M. Identification and assessment of known and novel human papillomaviruses by PCR amplification, restriction fragment length polymrophisms, nucleotide sequence, and phylogenetic algoritms. *J Infect Dis.* 1994;170:1077-1085.

Bosch, F.X., Castellsague, X., Munoz, N., Desanjose, S., Ghaffari, A.M., Gonzalez, L.C., Gili, M., Izarzugaza, I., Viladiu, P., Navarro, C., Vergara, A., Ascunce, N., Guerrero, E., and Shah, K.V. Male sexual behavior and human papillomavirus DNA - Key risk factors for cervical cancer in Spain. *J Natl Cancer Inst.* 1996;88(15):1060-1067.

Bosch, F.X., Munoz, N., Desanjose, S., Navarro, C., Moreo, P., Ascunce, N., Gonzalez, L.C., Tafur, L., Gili, M., Larranaga, I., Viladiu, P., Daniel, R.W., Deruiz, P.A., Aristizabal, N., Santamaria, M., Guerrero, E., and Shah, K.V. Human papillomavirus and cervical intraepithelial neoplasia grade-3 carcinoma insitu: a case-control study in Spain and Colombia. *Cancer Epidemiol Biomarkers Prev.* 1993;2:415-422.

Boshart, M., Gissmann, L., Ikenberg, H., Kleinheinz, A., Scheurlen, W., and zur Hausen, H. A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer. *EMBO Journal.* 1984;3:1151-1157.

Brinton, L.A. Epidemiology of cervical cancer - overview. In: Munoz, N., Bosch, F.X., Shah, K.V., and Meheus, A., (eds.). *The Epidemiology of Human Papillomavirus and Cervical Cancer.* Lyon: WHO-IARC; 1992:3-23.

Brinton, L.A. and Fraumeni, J.F.Jr. Epidemiology of uterine cervical cancer. *J Chronic Dis.* 1986;39:1051-1065.

Brinton, L.A., Hamman, R.F., Huggins, G.R., Lehman, H.F., Levine, R.S., Mallin, K., and Fraumeni, J.F.Jr. Sexual and reproductive risk factors for invasive squamous cell cervical cancer. *J Natl Cancer Inst.* 1987;79:23-30.

Brinton, L.A., Reeves, W.C., Brenes, M.M., Herrero, R., Gaitan, E., Tenorio, F., Britton, R.C., Garcia, M., and Rawls, W.E. The male factor in the etiology of cervical cancer among sexually monogamous women. *Int J Cancer.* 1989;44:199-203.

Brock, K.E., Berry, G., Brinton, L.A., Kerr, C., MacLennan, R., Mock, P.A., and Shearman, R.P. Sexual, reproductive and contraceptive risk factors for carcinoma-in-situ of the uterine cervix in Sydney. *Med J Aust.* 1989;150:125-130.

Chan, S.Y., Bernard, H.U., Ong, C.K., Chan, S.P., Hofmann, B., and Delius, H. Phylogenetic analysis of 48 papillomavirus types and 28 subtypes and variants: a showcase for the molecular evolution of DNA viruses. *J Virol.* 1992;66:5714-5725.

Chan, S.Y., Delius, H., Halpern, A.L., and Bernard H.U. Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny, and taxonomy. *J Virol.* 1995;69:3074-3083.

Chow, L.T., Hirochika, H., Nasseri, M., Stoler, M.H., Wolinsky, S.M., Chin, M.T., Hirochika, R., Arvan, D.S., and Broker, T.R. Human papillomavirus Gene Expression. In. *Cancer Cells 5 / Papillomaviruses.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory; 1987:55-72.

Coggin, J.R. and zur Hausen, H. Meeting Report - Workshop on papillomaviruses and cancer. *Cancer Res.* 1979;39:545-546.

Coleman, N., Birley, H.D., Renton, A.M., Hanna, N.F., Ryait, B.K., Byrne, M., Taylor-Robinson, D., and Stanley, M.A. Immunological events in regressing genital warts. *Am J Clin Pathol.* 1994;102:768-774.

Conrad, M., Bubb, V.J., and Schlegel, R. The human papillomavirus type 6 and 16 E5 proteins are membrane-associated proteins which associate with the 16-kilodalton pore-forming protein. *J Virol.* 1993;67:6170-6178.

Cox, D.R. Partial likelihood. *Biometrika.* 1975;62:269-276.

Cox, D.R. Regression models and life-tables. *J R Stat Soc.* 1972;34:187-220.

Cripe, T.P., Haugen, T.H., Turk, J.P., Tabatabai, F., Schmid, P.G.3., Durst, M., Gissmann, L., Roman, A., and Turek, LP. Transcriptional regulation of the human papillomavirus-16 E6-E7 promoter by a keratinocyte-dependent enhancer, and by viral E2 trans-activator and repressor gene products: implications for cervical carcinogenesis. *EMBO Journal.* 1987;6:3745-3753.

Critchlow, C.W. and Koutsky, L.A. Epidemiology of human papillomavirus infection. In: Mindel, A., (ed.). *Genital Warts. Human Papillomavirus Infection.* London, England: Edward Arnold; 1995:53-81.

Crook, T., Morgenstern, J.P., Crawford, L., and Banks, L. Continued expression of HPV-16 E7 protein is required for maintenance of the transformed phenotype of cells co-transformed by HPV-16 plus EJ-ras. *EMBO Journal.* 1989;8:513-519.

Crum, C.P., Barber, S., Symbula, M., Snyder, K., Saleh, A.M., and Roche, J.K. Coexpression of the human papillomavirus type 16 E4 and L1 open reading frames in early cervical neoplasia. *Virology.* 1990;178: 1:238-246.

Crum, C.P., Nuovo, G., Friedman, D., and Silverstein, S.J. Accumulation of RNA homologous to human papillomavirus type 16 open reading frames in genital precancers. *J Virol.* 1988;62: 1:84-90.

Dürst, M., Gissmann, L., Ikenberg, H., and zur Hausen, H. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. *Proc Natl Acad Sci U S A*. 1983;80:3812-3815.

Daniel, B., Rangarajan, A., Mukherjee, G., and Vallikad, E. The link between integration and expression of human papillomavirus type 16 genomes and cellular changes in the evolution of cervical intraepithelial neoplastic lesions. *J Gen Virol*. 1997;78:1095-1101.

de Sanjose, S., Bosch, F.X., Munoz, N., Tafur, L., Gili, M., Izarzugaza, I., Izquierdo, A., Navarro, C., Vergara, A., Munoz, M.T., Ascunce, N., and Shah, K.V. Socioeconomic differences in cervical cancer: two case-control studies in Colombia and Spain. *Am J Public Health*. 1996;86:1532-1538.

De Stavola, B. Statistical facts about cancers on which doctor Rigoni-Stern based his contribution to the surgeons' subgroup on the IV congress of the Italian scientists on 23 September 1842. *Stat Med*. 1987;6:881-884.

De Villiers, E.M. Human pathogenic papillomavirus types: An update. In: De Villiers, E.M., (ed.). *Current Topics in Microbiology and Immunology, Vol. 186*. Berlin - Heidelberg: Springer-Verlag; 1994:1-12.

Delius, H. and Hofmann, B. Primer-directed sequencing of human papillomavirus types. In: De Villiers, E.M., (ed.). *Current Topics in Microbiology and Immunology, Vol. 186*. Berlin - Heidelberg: Springer-Verlag; 1994:13-31.

DiLuca, D., Rotola, A., Pilotti, S., Monini, P., Caselli, E., Rilke, F., and Cassai, E. Simultaneous presence of herpes simplex virus and human papilloma virus sequences in human genital tumors. *Int J Cancer*. 1987;40:763-768.

Doorbar, J., Ely, S., Sterling, J., McLean, C., and Crawford, L. Specific interaction between HPV-16 E1-E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network. *Nature*. 1991;352:824-827.

Doorbar, J. and Gallimore, P.H. Identification of proteins encoded by the L1 and L2 open reading frames of human papillomavirus 1a. *J Virol*. 1987;61:2793-2799.

Dyson, N., Howley, P.M., Munger, K., and Harlow, E.D. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science*. 1989;243:934-937.

Evander, M., Edlund, K., Gustafsson, A., Jonsson, M., Karlsson, R., Rylander, E., and Wadell, G. Human papillomavirus infection is transient in young women - a population-based cohort study. *J Infect Dis*. 1995;171:1026-1030.

Fairley, C.K., Chen, S.J., Ugoni, A., Tabrizi, S.N., Forbes, A., and Garland, S.M. Human papillomavirus infection and its relationship to recent and distant sexual partners. *Obstet Gynecol.* 1994;84:755-759.

Fairley, C.K., Chen, S., Tabrizi, N., Leeton, K., Quinn, M.A., and Garland, S.M. The absence of genital human papillomavirus DNA in virginal women. *Int J STD AIDS.* 1992;3:414-417.

Favre, M., Breitburd, F., Croissant, O., and Orth, G. Structural polypeptides of rabbit, bovine, and human papillomaviruses. *J Virol.* 1975;15:1239-1247.

Franco, E.L. Association of HPV and anogenital cancer and implications for screening policy - Summary and discussion. In: Munoz, N., Bosch, F.X., Shah, K.V., and Meheus, A., (eds.). *The Epidemiology of Human Papillomavirus and Cervical Cancer.* Lyon: WHO-IARC; 1992:283-284.

Franco, E.L. Human papillomavirus and the natural history of cervical cancer. *Infect Med.* 1993;10:57-64.

Franco, E.L. Measurement errors in epidemiological studies of human papillomavirus and cervical cancer. In: Munoz, N., Bosch, F.X., Shah, K.V., and Meheus, A., (eds.). *The Epidemiology of Human Papillomavirus and Cervical Cancer.* Lyon: WHO-IARC; 1992:181-197.

Franco, E.L. The sexually transmitted disease model for cervical cancer: incoherent epidemiologic findings and the role of misclassification of human papillomavirus infection. *Epidemiology.* 1991;2:98-106.

Franco, E.L. Understanding the epidemiology of genital infection with oncogenic and non-oncogenic human papillomaviruses: A promising lead for primary prevention of cervical cancer. *Cancer Epidemiol Biomarkers Prev.* 1997;6:(in press).

Franco, E.L. Viral etiology of cervical cancer: a critique of the evidence. *Rev Infect Dis.* 1991;13:1195-1206.

Franco, E.L., Campos-Filho, N., Villa, L.L., and Torloni, H. Correlation patterns of cancer relative frequencies with some socioeconomic and demographic indicators in Brazil: an ecologic study. *Int J Cancer* . 1988;41:24-29.

Franco EL, Villa LL, Rahal P, Rohan TE. Incidence of cervical HPV infection in a cohort of women from a high-risk area for cervical cancer. 15th International Papillomavirus Workshop - Conference Proceedings: December 1, 1996-December 6, 1996; Gold Coast, Australia.

Franco, E.L., Villa, L.L., Ruiz, A., and Costa, M.C. Transmission of cervical human papillomavirus infection by sexual activity: differences between low and high oncogenic risk types. *J Infect Dis.* 1995;172:756-763.

Frazer, Y.H. and Tindle, R.W. Cell mediated immunity to papillomaviruses. In: Lacey, C., (ed.). *Papillomavirus Reviews: Current Research on Papillomaviruses.* London: Leeds University Press; 1996:151-163.

Galloway, D.A. HPV serology: another update. In: Lacey, C., (ed.). *Papillomavirus Reviews: Current Research on Papillomaviruses.* London: Leeds University Press; 1996:113-119.

Gissmann, L., Pfister, H., and zur Hausen, H. Human papilloma viruses (HPV): characterization of four different isolates. *Virology.* 1977;76:569-580.

Graham, S., Rawls, W.E., Swanson, M., and McCurtis, J. Sex partners and herpes simplex virus type 2 in the epidemiology of cancer of the cervix. *Am J Epidemiol.* 1982;115:729-735.

Gusberg, S.B. and Runowicz, C.D. Gynecologic Cancers. In: Holleb, A.L., Fink, D.J., and Murphy, G.P., (eds). *Textbook of Clinical Oncology.* Atlanta: American Cancer Society; 1991:481-485.

Gutman, L.T., Herman-Giddens, M.E., and Phelps, W.C. Transmission of human genital papillomavirus disease: comparison of data from adults and children. *Pediatrics.* 1993;91:31-38.

Gutman, L.T., St. Claire, K.K., Everett, V.D., Ingram, D.L., Soper, J., Johnston, W.W., Mulvaney, G.G., and Phelps, W.C. Cervical-vaginal and intraanal human papillomavirus infection of young girls with external genital warts. *J Infect Dis.* 1994;170:339-344.

Hagensee, M.E., Olson, N.H., Baker, T.S., and Galloway, D.A. Three-dimensional structure of vaccinia virus-produced human papillomavirus type 1 capsids. *J Virol.* 1994;68:4503-4505.

Hagey J, Gentleman J, Gaudette L. *Cancer in Canada 1991, Health Statistics Division, Statistics Canada.* Ottawa, Canada: Minister of Industry, Science and Technology; 1995.

Halbert, C.L., Demers, G.W., and Galloway, D.A. The E7 gene of human papillomavirus type 16 is sufficient for immortalization of human epithelial cells. *J Virol.* 1991;65:473-478.

Hawley-Nelson, P., Vousden, K.H., Hubbert, N.L., Lowy, D.R., and Schiller, J.T. HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. *EMBO Journal*. 1989;8:3905-3910.

Heinzl, S., Szalmay, G., Jochum, L., and Roemer, V. Observations on the development of dysplasia. *Acta Cytol*. 1982;26:453-456.

Herrero, R., Brinton, L.A., Reeves, W.C., Brenes, M.M., Tenorio, F., Debritton, R.C., Gaitan, E., Garcia, M., and Rawls, W.E. Sexual behavior, venereal diseases, hygiene practices, and invasive cervical cancer in a high-risk population. *Cancer*. 1990;65:380-386.

Herrero, R., Potischman, N., Brinton, L.A., Reeves, W.C., Brenes, M.M., Tenorio, F., Debritton, R.C., and Gaitan, E. A case-control study of nutrient status and invasive cervical cancer .I. dietary indicators. *Am J Epidemiol*. 1991;134:1335-1346.

Hildesheim, A., Gravitt, P., Schiffman, M.H., Kurman, R.J., Barnes, W., Jones, S., Tchabo, J.G., Brinton, L.A., Copeland, C., Epp, J., and Manos, M.M. Determinants of genital human papillomavirus infection in low-Income women in Washington, D.C. *Sex Transm Dis*. 1993;20:279-285.

Hildesheim, A., Schiffman, M.H., Gravitt, P.E., Glass, A.G., Greer, C.E., Zhang, T., Scott, D.R., Rush, B.B., Lawler, P., Sherman, M.E., and et al. Persistence of type-specific human papillomavirus infection among cytologically normal women. *J Infect Dis*. 1994;169:235-240.

Ho, L., Tay, S.K., Chan, S.Y., and Bernard, H.U. Sequence variants of human papillomavirus type 16 from couples suggest sexual transmission with low infectivity and polyclonality in genital neoplasia. *J Infect Dis*. 1993;168: 4:803-809.

Hoppe-Seyler, F. and Butz, K. Cellular control of human papillomavirus oncogene transcription. *Mol Carcinog*. 1994;10:134-141.

Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York: John Wiley & Sons; 1989.

Howley, P.M. Papillomavirinae: The viruses and their replication. In: Fields, B.N., Knipe, D.M., Howley, P.M., and et al., (eds.). *Fields Virology*. Philadelphia: Lippincott-Raven Publishers; 1996:2045-2076.

Husman, A.M.D.R., Walboomers, J.M.M., Vandenbrule, A.J.C., Meijer, C.J.L.M., and Snijders, P.J.F. The use of general primers GP5 and GP6 elongated at their 3' ends with adjacent highly conserved sequences improves human papillomavirus detection by PCR. *J Gen Virol*. 1995;76:1057-1062.

IARC Working Group. *Human Papillomaviruses. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 64.* Lyon: World Health Organization, International Agency for Research on Cancer; 1995.

Johnson, K. Periodic health examination, 1995 update: Screening for human papillomavirus infection in asymptomatic women. *Can Med Assoc J.* 1995;152:483-493.

Kaplan, E.L. and Meier, P. Nonparametric estimation from incomplete observations. *Am Stat Assoc J.* 1958;53:457-481.

Kessis, T.D., Slebos, R.J., Nelson, W.G., Kastan, M.B., Plunkett, B.S., Han, S.M., Lorincz, A.T., Hedrick, L., and Cho, K.R. Human papillomavirus 16 E6 expression disrupts the p53-mediated cellular response to DNA damage. *Proceedings of the National Academy of Sciences of the United States of America.* 1993;90:3988-92.

Kiviat, N.B., Koutsky, L.A., Critchlow, C.W., Galloway, D.A., Vernon, D.A., Peterson, M.L., McElhose, P.E., Pendras, S.J., Stevens, C.E., and Holmes, K.K. Comparison of Southern transfer hybridization and dot filter hybridization for detection of cervical human papillomavirus infection with types 6, 11, 16, 18, 31, 33, and 35. *Am J Clin Pathol.* 1990;94:561-565.

Kjaer, S.K., De Villiers, E.M., Caglayan, H., Svare, E., Haugaard, B.J., Engholm, G., Christensen, R.B., Moller, K.A., Poll, P., Jensen, H., Vestergaard, B.F., Lynge, E., and Jensen, O.M. Human papillomavirus, Herpes-simplex virus and other potential risk factors for cervical cancer in a high-risk area (Greenland) and a low-risk area (Denmark): A second look. *Br J Cancer.* 1993;67:830-837.

Kjaer, S.K., Devilliers, E.M., Dahl, C., Engholm, G., Bock, J.E., Vestergaard, B.F., Lynge, E., and Jensen, O.M. Case-control study of risk factors for cervical neoplasia in Denmark. 1: role of the male factor in women with one lifetime sexual partner. *Int J Cancer.* 1991;48:39-44.

Kjaer, S.K., Engholm, G., Teisen, C., Haugaard, B.J., Lynge, E., Christensen, R.B., Moller, K.A., Jensen, H., Poll, P., Vestergaard, B.F., DeVilliers, E.-M., and Jensen, O.M. Risk factors for cervical human papillomavirus and herpes simplex virus infections in Greenland and Denmark: a population-based study. *Am J Epidemiol.* 1990;131:669-682.

Kleinbaum DG. *Survival Analysis: a Self-Learning Text.* New York: Springer-Verlag; 1996.

Koutsky, L.A., Holmes, K.K., Critchlow, C.W., Stevens, C.E., Paavonen, J., Beckmann, A.M., DeRouen, T.A., Galloway, D.A., Vernon, D., and Kiviat, N.B. A cohort study of the risk of cervical intraepithelial neoplasia grade 2 or 3 in relation to papillomavirus infection. *N Engl J Med.* 1992;327:1272-1278.

Kunin, C.M. and Ames, R.E. Methods for determining the frequency of sexual intercourse and activities of daily living in young women. *Am J Epidemiol* 1981;113:55-61.

Laimins, L.A. The biology of human papillomaviruses: from warts to cancer. *Infect Agents Dis.* 1993;2:74-86.

Lambert, P.F., Pan, H., Pitot, H.C., Liem, A., Jackson, M., and Griep, A.E. Epidermal cancer associated with expression of human papillomavirus type 16 E6 and E7 oncogenes in the skin of transgenic mice. *Proc Natl Acad Sci U S A.* 1993;90:5583-5587.

Lane, D.P. Cancer. p53, guardian of the genome. *Nature.* 1992;358:15-16.

Last, J.M., (ed.). *A Dictionary of Epidemiology* . Third edition, New York: Oxford University Press; 1995.

LaVecchia, C., Franceschi, S., DeCarli, A., Fasoli, M., Gentile, A., Parazzini, F., and Regalo, M. Sexual factors, venereal diseases, and the risk of intraepithelial and invasive cervical cancer. *Cancer.* 1986;58:935-941.

LaVecchia, C., Franceschi, S., DeCarli, A., Fasoli, M., Gentile, A., and Tognoni, G. Cigarette smoking and the risk of cervical neoplasia. *Am J Epidemiol.* 1986;123:22-29.

Leechanachai, P., Banks, L., Moreau, F., and Matlashewski, G. The E5 gene from human papillomavirus type 16 is an oncogene which enhances growth factor-mediated signal transduction to the nucleus. *Oncogene.* 1992;7:19-25.

Ley, C., Bauer, H.M., Reingold, A., Schiffman, M.H., Chambers, J.C., Tashiro, C.J., and Manos, M.M. Determinants of genital human papillomavirus infection in young women. *J Natl Cancer Inst.* 1991;83:997-1003.

Li, J.Y., Li, F.P., Blot, W., Miller, R., and Fraumeni, J.J. Correlation between cancers of the uterine cervix and penis in China. *J Natl Cancer Inst.* 1982;69:1063-1065.

Lorincz, A.T., Reid, R., Jenson, A.B., Greenberg, M.D., Lancaster, W., and Kurman, R.J. Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types. *Obstet Gynecol* . 1992;79:328-337.

Münger, K., Werness, B.A., Dyson, N., Phelps, W.C., and Howley, P.M. Complex formation of human papillomavirus E7 proteins with the retinoblastoma tumour suppressor gene product. *EMBO Journal.* 1989;8:4099-4105.

Manos, M.M., Ting, Y., Wright, D.K., Lewis, A.J., Broker, T.R., and Wolinski, S.M. Use of polymerase chain reaction amplification for the detection of genital human papillomaviruses. *Cancer Cells.* 1989;7:209-214.

Mantel, N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*. 1966;50:163-170.

Matulic, M. and Soric, J. Papillomavirus genomes in human cervical carcinoma: analysis of their integration and transcriptional activity. *Neoplasma*. 1994;41:95-100.

McBride, A.A., Romanczuk, H., and Howley, P.M. The papillomavirus E2 regulatory proteins. *J Biol Chem*. 1991;266:18411-18414.

McLaughlin JR, Gentleman JF, Kaegi E, Levy I, Whittaker H, Steering Committee Members. *Canadian Cancer Statistics 1996*. Toronto, Canada: National Cancer Institute of Canada; 1996.

Melkert, P.W.J., Hopman, E., van den Brule, A.J.C., Risse, E.K.J., Vandiest, P.J., Bleker, O.P., Helmerhorst, T., Schipper, M.E.I., Meijer, C.J.L.M., and Walboomers, J.M.M. Prevalence of HPV in cytomorphologically normal cervical smears, as determined by the polymerase chain reaction, is age-dependent. *Int J Cancer*. 1993;53:919-923.

Mickey, R.M. and Greenland, S. The impact of confounder selection criteria on effect estimation. *Am J Epidemiol*. 1989;129:125-137.

Miller, A.B., Anderson, G., Brisson, J., Laidlaw, J., Lepitre, N., Malcolmson, P., Mirwaldt, P., Stuart, G., and Sullivan, W. Report of a National Workshop on Screening for Cancer of the Cervix. *Can Med Assoc J*. 1991;145:1301-1324.

Mirra, A.P. and Franco, E.L. Cancer incidence in Sao Paulo, Brazil. In: Franco, E.L., (ed.). *LICR Monograph Series in Cancer Epidemiology, Vol. 1*. Ludwig Institute for Cancer Research; 1985.

Mohr, I.J., Clark, R., Sun, S., Androphy, E.J., MacPherson, P., and Botchan, M.R. Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. *Science*. 1990;250:1694-1699.

Morrison, E.A., Ho, G.Y., Vermund, S.H., Goldberg, G.L., Kadish, A.S., Kelley, K.F., and Burk, R.D. Human papillomavirus infection and other risk factors for cervical neoplasia: a case-control study. *Int J Cancer*. 1991;49:6-13.

Moscicki, A.B., Palefsky, J., Smith, G., Siboshski, S., and Schoolnik, G. Variability of human papillomavirus DNA testing in a longitudinal cohort of young women. *Obstet Gynecol*. 1993;82:578-585.

Munoz, N. and Bosch, F.X. HPV and cervical neoplasia: review of case-control and cohort studies. In: Munoz, N., Bosch, F.X., Shah, K.V., and Meheus, A., (eds.). *The Epidemiology of Human Papillomavirus and Cervical Cancer*. Lyon: WHO-IARC; 1992:251-261.

Munoz, N., Castellsague, X., Bosch, F.X., Tafur, L., Desanjose, S., Aristizabal, N., Ghaffari, A.M., and Shah, K.V. Difficulty in elucidating the male role in cervical cancer in Colombia, a high-risk area for the disease. *J Natl Cancer Inst.* 1996;88:1068-1075.

Murphy, F.A., Fauquet, C.M., Bishop, D.H.L., Ghabrial, S.A., Jarvis, A.W., Martelli, G.P., Mayo, M.A., and Summers, M.D., (eds.). *Virus Taxonomy: Classification and Nomenclature of Viruses: Sixth Report of the International Committee on Taxonomy of Viruses.* Wien: Springer-Verlag; 1995.

Naguib, S.M., Lundin, F.E.J., and Davis, H.J. Relation of various epidemiologic factors to cervical cancer as determined by a screening program. *Obstet Gynecol.* 1966;28:4451-4459.

Nasiell, K., Roger, V., and Nasiell, M. Behavior of mild cervical dysplasia during long-term follow-up. *Obstet Gynecol.* 1986;67:665-669.

Nuovo, G.J. and Nuovo, J. Should family physicians test for human papillomavirus infection: an opposing view. *J Fam Practice.* 1991;32:188-192.

Orth, G., Jablonska, S., Favre, M., Croissant, O., Jarzabek-Chorzelska, M., and Rzesa, G. Characterization of two types of human papillomaviruses in lesions of epidermodysplasia verruciformis. *Proc Natl Acad Sci U S A.* 1978;75:1537-1541.

Palefsky, J.M. and Holly, E.A. Molecular virology and epidemiology of human papillomavirus and cervical cancer. *Cancer Epidemiol Biomarkers Prev.* 1995;4:415-428.

Pao, C.C., Tsai, P.L., Chang, Y.L., Hsieh, T.T., and Jin, J.Y. Possible non-sexual transmission of genital human papillomavirus infections in young women. *Eur J Clin Microbiol Infect Dis.* 1993;12:221-222.

Parkin, D.M., Pisani, P., and Ferlay, J. Estimates of the worldwide incidence of 18 major cancers in 1985. *Int J Cancer.* 1993;54:594-606.

Peters, R.K., Thomas, D., Hagan, D.G., Mack, T.M., and Henderson, B.E. Risk factors for invasive cervical cancer among latinas and non-latinas in Los Angeles County. *J Natl Cancer Inst.* 1986;77:1063-1077.

Peyton, C.L., Jansen, A.M., Wheeler, C.M., Stewart, A.C., Peto, J., Bosch, X.F., Munoz, N., Teyssie, A.R., Torroella, M., and Wabinga, H.R. A novel human papillomavirus sequence from an international cervical cancer study. *J Infect Dis.* 1994;170:1093-1095.

Peyton, C.L. and Wheeler, C.M. Identification of five novel human papillomavirus sequences in the New Mexico triethnic population. *J Infect Dis.* 1994;170:1089-1092.

Pim, D., Collins, M., and Banks, L. Human papillomavirus type 16 E5 gene stimulates the transforming activity of the epidermal growth factor receptor. *Oncogene*. 1992;7:27-32.

Prakash, S.S., Reeves, W.C., Sisson, G.R., Brenes, M., Godoy, J., Bacchetti, S., Britton, R.C., and Rawls, W.E. Herpes simplex virus type 2 and human papillomavirus type 16 in cervicitis, dysplasia and invasive cervical carcinoma. *Int J Cancer*. 1985;35:51-57.

Puranen, M., Yliskoski, M., Saarikoski, S., Syrjanen, K., and Syrjanen, S. Vertical transmission of human papillomavirus from infected mothers to their newborn babies and persistence of the virus in childhood. *Am J Obstet Gynecol*. 1996;174:694-699.

Reagan, J.W., Seidemann, I.L., Saracusa, Y. Cellular morphology of carcinoma *in situ* and dysplasia or atypical hyperplasia of uterine cervix. *Cancer*. 1953;6:224-225.

Reeves, W.C., Brinton, L.A., Garcia, M., Brenes, M.M., Herrero, R., Gaitan, E., Tenorio, F., Britton, R.C., and Rawls, W.E. Human papillomavirus infection and cervical cancer in Latin America. *N Engl J Med*. 1989;320:1437-1441.

Reid, R. and Lorincz, A.T. Should family physicians test for human papillomavirus infection? An affirmative view. *J Fam Practice*. 1991;32:183-188.

Resnick, R.M., Cornelissen, M.T., Wright, D.K., Eichinger, G.H., Fox, H.S., Schegget, J.T., and Manos, M.M. Detection and typing of human papillomavirus in archival cervical cancer specimens by DNA amplification with consensus primers. *J Natl Cancer Inst*. 1990;82:1477-1484.

Richardson H. *Risk Factors for Cervical HPV Infection in Montreal University Students*. Laval: Institut Armand-Frappier, Université du Québec; 1996. M.Sc. thesis.

Richart, R.M. Colpomicroscopic studies of cervical intraepithelial neoplasia. *Cancer*. 1966;19:395-405.

Richart, R.M. A theory of cervical carcinogenesis. *Obstetrical & Gynecological Survey*. 1969;24:874-879.

Richart, R.M. Screening. The next century. *Cancer*. 1995;76(10 Suppl):1919-1927.

Richart, R.M. and Barron, B.A. A follow-up study of patients with cervical dysplasia. *Am J Obstet Gynecol*. 1969;105:386-393.

Roden, R.B., Kirnbauer, R., Jenson, A.B., Lowy, D.R., and Schiller, J.T. Interaction of papillomaviruses with the cell surface. *J Virol*. 1994;68:7260-7266.

Rohan, T., Mann, V., McLaughlin, J., Harnish, D.G., Yu, H., Smith, D., Davis, R., Shier, R.M., and Rawls, W. PCR-detected genital papillomavirus infection: prevalence and association with risk factors for cervical cancer. *Int J Cancer*. 1991;49:1-5.

Rohan, T.E., McLaughlin, J.R., Harnish, D.G. Repeatability of interview-derived information on sexual history: a study in women. *Epidemiology*. 1994;5:360-363.

Romanczuk, H. and Howley, P.M. Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. *Proc Natl Acad Sci U S A*. 1992;89:3159-3163.

Romanczuk, H., Thierry, F., and Howley, P.M. Mutational analysis of cis elements involved in E2 modulation of human papillomavirus type 16 P97 and type 18 P105 promoters. *J Virol*. 1990;64:2849-2859.

Rothman KJ. *Modern Epidemiology*. Boston: Little, Brown and Company; 1986.

Rylander, E., Ruusuvaara, L., Almstromer, M.W., Evander, M., and Wadell, G. The absence of vaginal human papillomavirus 16 DNA in women who have not experienced sexual intercourse. *Obstet Gynecol*. 1994;83:735-737.

Sang, B.C. and Barbosa, M.S. Increased E6/E7 transcription in HPV 18-immortalized human keratinocytes results from inactivation of E2 and additional cellular events. *Virology*. 1992;189:448-455.

Scheffner, M., Romanczuk, H., Münger, K., Huibregtse, J.M., Mietz, J.A., and Howley, P.M. Functions of human papillomavirus proteins. In: zur Hausen, H., (ed.). *Current Topics in Microbiology and Immunology, Vol. 186*. Berlin - Heidelberg: Springer-Verlag; 1994:83-99.

Scheffner, M., Werness, B.A., Huibregtse, J.M., Levine, A.J., and Howley, P.M. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell*. 1990;63:1129-1136.

Schiffman, M.H. Epidemiology of cervical human papillomavirus infections. In: zur Hausen, H., (ed.). *Current Topics in Microbiology and Immunology, Vol. 186*. Berlin - Heidelberg: Springer-Verlag; 1994:55-81.

Schiffman, M.H. Validation of hybridization assays: correlation of filter in situ, dot blot and PCR with Southern blot. In: Munoz, N., Bosch, F.X., Shah, K.V., and Meheus, A., (eds.). *The Epidemiology of Human Papillomavirus and Cervical Cancer*. Lyon: WHO-IARC; 1992:169-179.

Schiffman, M.H., Bauer, H.M., Hoover, R.N., Glass, A.G., Cadell, D.M., Rush, B.B., Scott, D.R., Sherman, M.E., Kurman, R.J., Wacholder, S., Stanton, C.K., and Manos, M.M.

Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. *J Natl Cancer Inst*. 1993;85:958-964.

Schiffman, M.H., Haley, N.J., Felton, J.S., Andrews, A.W., Kaslow, R.A., Lancaster, W.D., Kurman, R.J., Brinton, L.A., Lannom, L.B., and Hoffmann, D. Biochemical epidemiology of cervical neoplasia: measuring cigarette smoke constituents in the cervix. *Cancer Res*. 1987;47:3886-3888.

Schiffman, M.H. and Schatzkin, A. Test reliability is critically important to molecular epidemiology: an example from studies of human papillomavirus infection and cervical neoplasia. *Cancer Res*. 1994;54:S1944-S1947.

Schiller, J.T., Vass, W.C., and Lowy, D.R. Identification of a second transforming region in bovine papillomavirus DNA. *Proc Natl Acad Sci U S A*. 1984;81:7880-7884.

Schneider, A. Genital HPV infections. In: Lacey, C., (ed.). *Papillomavirus Reviews: Current Research on Papillomaviruses*. London: Leeds University Press; 1996:207-214.

Schneider, A. and Koutsky, L.A. Natural history and epidemiological features of genital HPV infection. In: Munoz, N., Bosch, F.X., Shah, K.V., and Meheus, A., (eds.). *The Epidemiology of Human Papillomavirus and Cervical Cancer*. Lyon: World Health Organization, International Agency for Research on Cancer; 1992:25-52.

Schneider, A., Oltersdorf, T., Schneider, V., and Gissmann, L. Distribution pattern of human papilloma virus 16 genome in cervical neoplasia by molecular in situ hybridization of tissue sections. *International Journal of Cancer*. 1987;39:717-21.

Sedman, S.A., Barbosa, M.S., Vass, W.C., Hubbert, N.L., Haas, J.A., Lowy, D.R., and Schiller, J.T. The full-length E6 protein of human papillomavirus type 16 has transforming and trans-activating activities and cooperates with E7 to immortalize keratinocytes in culture. *J Virol*. 1991;65:4860-4866.

Shah, K.V. and Howley, P.M. Papillomaviruses. In: Fields, B.N., Knipe, D.M., Howley, P.M., and et al., (eds.). *Fields Virology*. Philadelphia: Lippincott-Raven Publishers; 1996:2077-2109.

Shamanin, V., Glover, M., Rausch, C., Proby, C., Leigh, I.M., zur Hausen, H., and De Villiers, E.M. Specific types of human papillomavirus found in benign proliferations and carcinomas of the skin in immunosuppressed patients. *Cancer Res*. 1994;54:4610-4613.

Shaw, P., Bovey, R., Tardy, S., Sahli, R., Sordat, B., and Costa, J. Induction of apoptosis by wild-type p53 in a human colon tumor-derived cell line. *Proc Natl Acad Sci U S A*. 1992;89:4495-4499.

Shingleton, H.M., Richart, R.M., Wiener, J., and Spiro, D. Human cervical intraepithelial neoplasia: fine structure of dysplasia and carcinoma in situ. *Cancer Res.* 1968;28:695-706.

Skinner, G.R. Transformation of primary hamster embryo fibroblasts by type 2 simplex virus: evidence for a "hit and run" mechanism. *British Journal of Experimental Pathology.* 1976;57:361-376.

Slattery, M.L., Overall, J.C., Abbott, T.M., French, T.K., Robinson, L.M., and Gardner, J.W. Sexual activity, contraception, genital infections, and cervical cancer: support for a sexually transmitted disease hypothesis. *Am J Epidemiol.* 1989;130:248-258.

Smotkin, D., Prokoph, H., and Wettstein, F.O. Oncogenic and nononcogenic human genital papillomaviruses generate the E7 mRNA by different mechanisms. *J Virol.* 1989;63:1441-1447.

Smotkin, D. and Wettstein, F.O. The major human papillomavirus protein in cervical cancers is a cytoplasmic phosphoprotein. *J Virol.* 1987;61:1686-1689.

Solomon, D. The 1988 Bethesda system for reporting cervical/vaginal cytologic diagnoses. Developed and approved at the National Cancer Institute Workshop, Bethesda, Maryland, USA, December, 12-13, 1988. *J Clin Cytol Cytopathol.* 1989;33:567-574.

SPSS inc. *SPSS Advanced Statistics 7.5.* Prentice-Hall; 1996.

Stern, E. and Neely, P.M. Dysplasia of the uterine cervix: incidence of regression, recurrence, and cancer. *Cancer.* 1964;17:508-512.

Straight, S.W., Herman, B., and McCance, D.J. The E5 oncoprotein of human papillomavirus type 16 inhibits the acidification of endosomes in human keratinocytes. *J Virol.* 1995;69:3185-3192.

Thierry, F., Carranca, A.G., and Yaniv, M. Elements that control the transcription of genital human papillomavirus type 18. In. *Cancer Cells 5 / Papillomaviruses.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory; 1987:23-32.

Thomas, D.B. An epidemiologic study of carcinoma in situ and squamous dysplasia of the uterine cervix. *Am J Epidemiol.* 1973;98:10-28.

Tommasino, M., Adamczewski, J.P., Carlotti, F., Barth, C.F., Manetti, R., Contorni, M., Cavalieri, F., Hunt, T., and Crawford, L. HPV16 E7 protein associates with the protein kinase p33CDK2 and cyclin A. *Oncogene.* 1993;8:195-202.

Ustav, M., Ustav, E., Szymanski, P., and Stenlund, A. Identification of the origin of replication of bovine papillomavirus and characterization of the viral origin recognition factor E1. *EMBO Journal*. 1991;10:4321-4329.

van den Brule, A.J.C., Snijders, P.J.F., Gordijn, R.L.J., Bleker, O.P., Meijer, C.J.L.M., and Walboomers, J.M.M. General primer-mediated polymerase chain reaction permits the detection of sequenced and still unsequenced human papillomavirus genotypes in cervical scrapes and carcinomas. *Int J Cancer*. 1990;45:644-649.

van den Brule, A.J.C., Walboomers, J.M.M., Dumaine, M., Kenemans, P., and Meijer, C.J.L.M. Difference in prevalence of human papillomavirus genotypes in cytomorphologically normal cervical smears is associated with a history of cervical intraepithelial neoplasia. *Int J Cancer*. 1991;48:404-408.

van Doornum, G.J.J., Prins, M., Juffermans, L.H.J., Hooykaas, C., van den Hoek, J.A.R., Coutinho, R.A., and Quint, W.G.V. Regional distribution and incidence of human papillomavirus infections among heterosexual men and women with multiple sexual partners: a prospective study. *Genitourin Med*. 1994;70:240-246.

van Doornum, G.J., Hooykaas, C., Juffermans, L.H.J., van der Lans, S.M.G.A., van der Linden, M.M.D., Coutinho, R.A., and Quint, W.G. Prevalence of human papillomavirus infections among heterosexual men and women with multiple sexual partners. *J Med Virol*. 1992;37:13-21.

VanEenwyk, J., Davis, F.G., and Colman, N. Folate, vitamin C, and cervical intraepithelial neoplasia. *Cancer Epidemiol Biomarkers Prev*. 1992;1:119-124.

Velu, T.J., Beguinot, L., Vass, W.C., Willingham, M.C., Merlino, G.T., Pastan, I., and Lowy, D.R. Epidermal-growth-factor-dependent transformation by a human EGF receptor proto-oncogene. *Science*. 1987;238:1408-1410.

Villa, L.L. and Franco, E.L. Epidemiologic correlates of cervical neoplasia and risk of human papillomavirus infection in asymptomatic women in Brazil. *J Natl Cancer Inst*. 1989;81:332-340.

von Knebel Doeberitz, M., Rittmuller, C., zur Hausen, H., and Durst, M. Inhibition of tumorigenicity of cervical cancer cells in nude mice by HPV E6-E7 anti-sense RNA. *Int J Cancer*. 1992;51:831-834.

von Krogh, G. Genitoanal papillomavirus infection: diagnostic and therapeutic objectives in the light of current epidemiological observations. *Int J STD & AIDS*. 1991;2:391-404.

Vonka, V., Kanka, J., and Roth, Z. Herpes simplex type 2 virus and cervical neoplasia. *Adv Cancer Res*. 1987;48:149-191.

Walker AM. *Observation and Inference: an Introduction to Epidemiology*. Newton Lower Falls, MA, USA: Epidemiology Ressources Inc.; 1991.

Werness, B.A., Levine, A.J., and Howley, P.M. Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science*. 1990;248:76-79.

Wheeler, C.M., Parmenter, C.A., Hunt, W.C., Becker, T.M., Greer, C.E., Hildesheim, A., and Manos, M.M. Determinants of genital human papillomavirus infection among cytologically normal women attending the University-of-New-Mexico student health center. *Sex Transm Dis*. 1993;20:286-289.

Wikström, A., van Doornum, G.J.J., Quint, W.G.V., Schiller, J.T., and Dillner, J. Identification of human papillomavirus seroconversions. *J Gen Virol*. 1995;76:529-539.

Wilson, V.G. and Ludes-Meyers, J. A bovine papillomavirus E1-related protein binds specifically to bovine papillomavirus DNA. *J Virol*. 1991;65:5314-5322.

Wright, T.C. and Kurman, R.J. A critical review of the morphologic classification systems of preinvasive lesions of the cervix: the scientific basis for shifting the paradigm. In: Lacey, C., (ed.). *Papillomavirus Reviews: Current Research on Papillomaviruses*. London: Leeds University Press; 1996:215-225.

Wright, T.C., Kurman, R.J., and Ferenczy A. Precancerous lesions of the cervix. In: Kurman, R.J., (ed). *Blaustein's Pathology of Female Genital Tract*. New York: Springer-Verlag; 1994:229-277.

Wynder, E.L., Cornfield, J., Shroff, P.D., and Doraiswami, K.R. A study of environmental factors in carcinoma of the cervix. *Am J Obstet Gynecol*. 1954;68:1016-1052.

Ziegler, R.G., Jones, C.J., Brinton, L.A., Norman, S.A., Mallin, K., Levine, R.S., Lehman, H.F., Hamman, R.F., Trumble, A.C., Rosenthal, J.F., and Hoover, R.N. Diet and the risk of in situ cervical cancer Among white women in the United States. *Cancer Causes Control*. 1991;2:17-29.

zur Hausen, H. Condylomata acuminata and human genital cancer. *Cancer Res*. 1976;36:794.

zur Hausen, H. Papillomaviruses in anogenital cancer as a model to understand the role of viruses in human cancers. *Cancer Res* . 1989;49:4677-4681.

zur Hausen, H. Preface. In: zur Hausen, H. *Current Topics in Microbiology and Immunology, Vol. 186*. Berlin - Heidelberg: Springer-Verlag; 1994:I-IX.

zur Hausen, H., Schulte-Holthausen, H., Wolf, H., Dörries, K., and Egger, H. Attempts to detect virus-specific DNA in human tumors. - II. Nucleic acid hybridizations with complementary RNA of human herpes groups viruses. *Int J Cancer*. 1974;13:657-64.

APPENDIX 1

BASELINE QUESTIONNAIRE

Data da entrevista: D: _____ M: _____ A: _____ Hora do início: _____

1. Nº no estudo: _____

2. Registro M.E.V.N.C: _____

3. Qual o seu nome? _____

4. Em que dia, mês e ano a Sra. nasceu? D: _____ M: _____ A: _____

   Portanto, quantos anos a Sra. tem? _____ anos

5. Grupo étnico (*Interpretação da entrevistadora*):
   ¦1¦ Branca ¦2¦ Mulata ¦3¦ Negra ¦4¦ Amarela ¦5¦ Índio ou descendente

6. Qual é o seu estado civil?
   ¦1¦ Solteira ¦2¦ Casada ¦3¦ Viúva
   ¦4¦ Separada ¦5¦ Vive maritalmente ("juntada" ou amigada)

7. Quais foram as suas ocupações/empregos nos últimos dez anos?

   _____

   _____

   _____

   _____

8. Até que grau escolar a Sra. estudou?
   ¦1¦ Analfabeta ¦2¦ Primário incompleto ¦3¦ Primário completo
   ¦4¦ Secundário incompleto ¦5¦ Secundário completo
   ¦6¦ Técnico-profissionalizante ¦7¦ Superior

9. Qual a sua religião?
   ¦1¦ Católica ¦2¦ Crente ¦3¦ Protestante ¦4¦ Judia (Israelita)

   ¦5¦ Espírita ¦6¦ Umbandista ¦7¦ outra (qual?) _____ ¦8¦ não tem

10. Incluindo a Sra., quantas pessoas vivem na sua casa? _____ pessoas

11. Qual é a sua renda familiar, ou seja, a da Sra. mais a dos que vivem em sua casa?

    CR$ _____ cruzeiros reais

12. Quais dos seguintes itens a Sra. tem em casa?

a) geladeira	¦1¦ Sim	¦2¦ Não
b) T.V. a cores	¦1¦ Sim	¦2¦ Não
c) telefone	¦1¦ Sim	¦2¦ Não
d) video-cassete	¦1¦ Sim	¦2¦ Não
e) carro	¦1¦ Sim	¦2¦ Não
f) carros adicionais	¦1¦ Sim	¦2¦ Não

13. Em que bairro a Sra. mora: _____

14. Há quantos anos a Sra. mora nesse local? _____ anos

15. Onde a Sra. nasceu?: Cidade:_____ Estado _____

16. Essa cidade era: ¦1¦ área rural ¦2¦ área urbana ¦3¦ subúrbio ¦8¦ não sabe

17. Onde a Sra. morou a maior parte de sua vida?
(*após os 12 anos de idade*) Cidade: _____ Estado _____

18. Esse local era: ¦1¦ área rural ¦2¦ área urbana ¦3¦ subúrbio ¦8¦ não sabe

19. A Sra. fuma ou já fumou?
¦1¦ Sim ¦2¦ Nunca (*se nunca, vá direto à questão 31*)

Se sim, fazer as perguntas referentes a cada tipo de tabaco:

Cigarros de papel industrializados

20. Quantos cigarros a Sra. fuma/fumava em média por dia, aproximadamente?
¦1¦ no máximo 1 ¦2¦ de 2 a 5 ¦3¦ de 6 a 10 ¦4¦ 11 a 20
¦5¦ mais que 20 ¦6¦ mais que 40 (2 maços)

21. Que tipos de cigarro a Sra. fuma/fumava?
¦1¦ somente com filtro ¦2¦ principalmente com filtro, as vezes sem filtro
¦3¦ principalmente sem filtro, as vezes com filtro ¦4¦ somente sem filtro

22. Com que idade a Sra. começou a fumar regularmente ? _____ anos

23. (*Se ainda fuma, perguntar*) Há quantos anos a Sra. fuma? _____ anos

24. (*Se parou, perguntar*) Durante quantos anos a Sra. fumou? _____ anos

Cigarros de fumo de corda, de palha ou papel

25. Quantos cigarros de fumo de corda ou palha a Sra. fuma/fumava em média por dia, aproximadamente?
¦1¦ no máximo 1 ¦2¦ de 2 a 5 ¦3¦ de 6 a 10 ¦4¦ 11 a 20
¦5¦ mais que 20 ¦6¦ mais que 40 (2 maços)

26. Com que idade a Sra. começou a fumar regularmente? _____ anos

27. (*Se ainda fuma, perguntar*) Há quantos anos a Sra. fuma? _____ anos

28. (*Se parou, perguntar*) Durante quantos anos a Sra. fumou? _____ anos

29. (*Se parou, perguntar*) Há quantos anos a Sra. parou? _____ anos

30. A Sra. fuma/fumou charuto ou cachimbo? ¦1¦ Sim ¦2¦ Não

Consumo de bebidas alcoólicas

31. A Sra. costuma/costumava consumir bebidas alcoólicas, mesmo que ocasionalmente?
¦1¦ Sim ¦2¦ Nunca *(se nunca, vá direto à q. 38)*

32. A Sra. costuma/costumava beber cerveja?
¦1¦ Não/ocasionalmente ¦2¦ no máximo um copo por semana
¦3¦ de 2 a 5 por semana ¦4¦ de 6 a 10 ¦5¦ 11 a 30 ¦6¦ mais que 30

33. A Sra. costuma/costumava beber vinho?
¦1¦ Não/ocasionalmente ¦2¦ no máximo um copo por semana
¦3¦ de 2 a 5 por semana ¦4¦ de 6 a 10 ¦5¦ 11 a 30 ¦6¦ mais que 30

34. A Sra. costuma/costumava beber pinga ou cachaça?
¦1¦ Não/ocasionalmente ¦2¦ no máximo um copo por semana
¦3¦ de 2 a 5 por semana ¦4¦ de 6 a 10 ¦5¦ 11 a 30 ¦6¦ mais que 30

35. A Sra. costuma/costumava uisque, gim, vodca ou outra bebida forte?
¦1¦ Não/ocasionalmente ¦2¦ no máximo um copo por semana
¦3¦ de 2 a 5 por semana ¦4¦ de 6 a 10 ¦5¦ 11 a 30 ¦6¦ mais que 30

36. Há quantos anos a Sra. bebe essas quantidades? _____ anos
(as referidas acima?)

37. Durante quantos anos a Sra. bebeu? _____ anos
(Se parou de beber, perguntar)

Eu gostaria agora de lhe fazer algumas perguntas sobre sua vida íntima. Eu entendo que este é um assunto pessoal, mas conhecer estas informações será de grande auxílio na nossa pesquisa. Eu volto a lembrar a Sra. que todas as respostas serão mantidas em total segredo. Nunca estes dados serão revelados a alguém.

38. Que idade a Sra. tinha quando menstruou pela primeira vez? _____ anos
(se menopausada e passar a q. 11)

39. Quando a Sra. teve a sua última menstruação? D: ___ M: ___ A: ___

¦_¦ Puérpera ¦_¦ Lactante ¦_¦ Menopausada

40. Quando está/estava menstruada, o que a Sra. usa/usava como absorvente íntimo?

a) absorvente tipo "MODESS" comercial	¦1¦ Sim	¦2¦ Não
b) absorvente interno tipo OB/Tampax	¦1¦ Sim	¦2¦ Não
c) toalhinha de pano	¦1¦ Sim	¦2¦ Não
d) outro (*qual?*_____)	¦1¦ Sim	¦2¦ Não

41. Nos últimos cinco anos, quantas vezes a Sra. sentiu coceira na região genital?
¦1¦ Nenhuma vez ¦2¦ Algumas vezes (1-9) ¦3¦ Muitas vezes (10+)

42. Nos últimos cinco anos, quantas vezes a Sra. sentiu dor/ardor na região genital?
¦1¦ Nenhuma vez ¦2¦ Algumas vezes (1-9) ¦3¦ Muitas vezes (10+)

43. Nos últimos cinco anos, quantas vezes a Sra. teve corrimento vaginal?
¦1¦ Nenhuma vez ¦2¦ Algumas vezes (1-9) ¦3¦ Muitas vezes (10+)

44. Já fez ou faz uso de algum produto para tratamento ginecológico?
(mostrar ou ler a lista de nomes de medicamentos impressa no verso da página anterior)

_____

_____

45. A Sra. já utilizou algum produto que não seja de farmácia para tratamento ginecológico?
¦1¦ Sim (*qual:* _____) ¦2¦ Não

46. Nos últimos dois dias, a Sra. teve corrimento, coceira ou ardor na região genital? ¦1¦ Sim ¦2¦ Não

47a. A Sra. usa/já usou algum sistema que force água/líquidos para o interior da vagina tais como duchas, bidês, etc?
¦1¦ sim, sempre ¦2¦ sim, frequentemente
¦3¦ de vez em quando ¦4¦ nunca

47b. (*Se sim*) Com que produto? _____

48. Durante os períodos menstruais a Sra. costuma/costumava lavar seus órgãos genitais? (*Além do banho diário*)
¦1¦ Não ¦2¦ Sim, uma vez por dia ¦3¦ Sim, mais de uma vez por dia

49. A Sra. já teve alguma vez feridas na vagina ou vulva? ¦1¦ Sim ¦2¦ Não

50. Alguma vez a Sra. soube por um(a) médico(a) que tinha uma doença venérea ou sexualmente transmissível? (*se sim, qual?*)
¦1¦ Gonorreia ¦2¦ Cancro ¦3¦ Crista de galo ou condiloma ¦4¦ Sífilis
¦5¦ Herpes ¦6¦ Tricomoníase ¦7¦ Candidíase 8 Nunca teve

51. A Sra. já fez um exame de prevenção do câncer de colo de útero, também chamado Papanicolaou ou citológico ou citologia oncótica? ¦1¦ Sim ¦2¦ Não

52. (*Se sim*) Quantas vezes? _____ vezes

53. (*Se sim*) Quando foi a última vez que fez este exame?
¦1¦ No último ano ¦2¦ Há mais de um ano, mas menos que cinco
¦3¦ Há mais que cinco anos ¦8¦ Não lembra

Lembrete: Foro íntimo e confidencialidade

54. Com que idade a Sra. teve a sua primeira relação sexual? _____ anos
(Se virgem, vá direto a q. 100, depois de certificar-se que ela nunca engravidou)

55. Quantas vezes a Sra. já engravidou? _____ vezes
(Se nunca, vá direto a q. 63)

56. Quantas destas gestações resultaram em partos normais? _____

57. Quantas foram por operação cesariana? _____

58. Quantas resultaram em aborto? _____

59. Em que ano foi a sua última gravidez? _____

60. Foi uma gestação completa? ¦1¦ Sim ¦2¦ Não

61. Enquanto grávida a Sra. continuava tendo relações sexuais com seu marido/parceiro? ¦1¦ Sim ¦2¦ Não

62. A Sra. costuma/costumava resguardar-se de relações sexuais após cada parto? ¦1¦ Sim ¦2¦ Não

63. Com que idade a Sra. começou a ter relações sexuais pelo menos uma vez por mês

Aos _____ anos ->¦_¦ se nunca foi constante

64. Durante a sua vida inteira, com quantos homens a Sra. manteve relações sexuais?

(*Insista para que ela dê uma resposta mesmo q. aproximada*) _____

65. Quantos destes parceiros foram regulares, isto é, com os quais a Sra. teve relações sexuais regulares durante um periodo mínimo de 6 meses, independentemente de morar na mesma casa?

_____

66. Que a Sra. saiba, quantos destes parceiros sexuais não foram fiéis, isto é, tiveram contato sexual com outras mulheres?

_____

67. No total, quantos parceiros sexuais a Sra. teve antes de 20 anos? _____

68. Quantos destes parceiros (*antes dos 20*) tinham menos que 20 anos? _____

69. Quantos destes parceiros (*antes dos 20*) tinham mais que 30 anos? _____

Se a paciente tiver menos que 20 anos, passar para a q. 73

70. No total, quantos parceiros sexuais a Sra. teve depois dos 20 anos? _____

71. Quantos destes parceiros (*depois dos 20*) tinham menos que 20 anos? _____

72. Quantos destes parceiros (*depois dos 20*) tinham mais que 30 anos? _____

73. Desde o inicio de sua vida sexual houve periodos em que a Sra. não teve relações por mais que um ano? Se sim, quantos periodos *(total em anos)*?

_____ anos

74. Em geral, considerando a maior parte de sua vida sexual, com que frequência a Sra. tem mantido/manteve relações sexuais? *(Descrever a frequência e duração para cada periodo lembrado pela paciente)*

_____

_____

_____

75. Enquanto menstruada, a Sra. evita/evitava ter relações sexuais com seu marido/parceiro? (*Também considerar a hipótese do marido evitar)*
¦1¦ Sempre evitei ou marido evitou ¦2¦ De vez em quando
¦3¦ Só nos primeiros dias ¦4¦ Nunca

76. A Sra. costuma/costumava lavar seus genitais antes de ter relações sexuais?
¦1¦ Sempre ¦2¦ de vez em quando ¦3¦ Nunca

77. A Sra. costuma/costumava lavar seus genitais depois das relações sexuais?
¦1¦ Sempre ¦2¦ de vez em quando ¦3¦ Nunca

78. Durante os últimos cinco anos, com quantos homens a Sra. manteve relações sexuais? (*Insista para que ela dê uma resposta mesmo que aproximada*

_____

79. Que a Sra. saiba, quantos destes parceiros sexuais não foram fiéis, isto é, tiveram contato sexual com outras mulheres? _____

80. Quantos destes parceiros tinham menos que 20 anos? _____

81. Quantos destes parceiros tinham mais que 30 anos? _____

82. Durante este periodo dos últimos 5 anos, com que frequência a Sra. teve relações sexuais com seu marido ou parceiro(s)?

_____/semana _____/mes _____/ano

Dizer à paciente para se lembrar apenas dos últimos 12 meses

83. Durante os últimos 12 meses, com quantos homens a Sra. teve relações sexuais? (*Insista para que ela dê uma resposta mesmo que aproximada*)

_____

84. Que a Sra. saiba, quantos destes parceiros sexuais não foram fiéis, isto é, tiveram contato sexual com outras mulheres?

_____

85. Quantos destes parceiros tinham menos que 20 anos? _____

86. Quantos destes parceiros tinham mais que 30 anos? _____

87. Durante este periodo dos últimos 12 meses, com que frequência a Sra. teve relações sexuais com seu marido ou parceiro(s)?

_____/semana _____/mes _____/ano

Fim das perguntas ciclicas

Detalhes sobre os métodos anticoncepcionais usados

88. Quais são os métodos que a Sra. ou o seu marido/parceiro(s) tem usado/usaram para evitar filhos? (*assinalar todos os mencionados)*
¦1¦ pílula anticoncepcional ¦2¦ laqueadura ¦3¦ vasectomia
¦4¦ D.I.U. ¦5¦ condom ¦6¦ diafragma
¦7¦ Geléia espermicida ¦8¦ Coito interrompido/tabelinha/muco cervical
¦9¦ Outro: _____ ¦10¦ Não sabe

89. *(Se P.A., perguntar)* Com que idade a Sra. começou a usar P.A.? _____ anos

90. *(Se P.A.)* Durante quantos anos a Sra. tem tomado/tomou P.A.? _____ anos

91. *(Se P.A.)* Durante este(s) período(s) a Sra. obedeceu os intervalos regulares de descanso recomendados pelo médico? ¦1¦ Sim ¦2¦ Não

92. *(Se parou, perguntar)* Há quanto tempo parou de tomar P.A.? _____ anos

93. *(Se laqueadura, perguntar)*
Há quanto tempo foi a laqueadura que a Sra. fez? _____ anos

94. *(Se vasectomia do parceiro mais frequente, perguntar)*
Há quanto tempo foi a vasectomia do seu marido/parceiro? _____ anos

95. *(Se D.I.U., perguntar)*
Com que idade a Sra. usou D.I.U. pela primeira vez? _____ anos

96. *(Se D.I.U.)* A Sra. ainda usa D.I.U.? ¦1¦ Sim ¦2¦ Não

97. *(Se condom, perguntar)* Com que frequência seu marido/parceiro(s) usa(m) camisinha? ¦1¦ Muito raramente ¦2¦ As vezes ¦3¦ Sempre

98. *(Se diafragma, perguntar)* com que frequência a Sra. tem utilizado/utilizou diafragma? ¦1¦ Muito raramente ¦2¦ As vezes ¦3¦ Sempre

99. *(Se geléia, perguntar)* A Sra. tem usado a geléia espermicida de que maneira?
¦1¦ Principalmente como método único ¦2¦Principal/ associado ao diafragma
¦3¦ Principalmente associado à camisinha

Lembrete: Foro intimo e confidencialidade

100. A Sra. já praticou/pratica coito anal, isto é, relação com penetração pelo anus?
¦1¦ Sim, frequentemente ¦2¦ Sim, raramente ¦3¦ Não

Se não, vá direto à questão 105

101. (*Se sim, perguntar*)
Com quantos parceiros a Sra. já praticou/pratica coito anal? _____

Se mais de 1 na resposta anterior, iniciar as próximas perguntas enfatizando que a entrevistada deve se referir ao parceiro com qu ela mais frequentemente praticou coito anal.

102. O seu marido/parceiro realizava/realiza penetração vaginal em seguida ao coito anal? ¦1¦ Sim ¦2¦ Não ¦3¦ As vezes

103. (*Se sim, perguntar*) Antes da penetração vaginal o seu marido/parceiro fazia/faz a higiene do pênis? ¦1¦ Sim ¦2¦ Não ¦3¦ As vezes

104. (*Alternativamente*) Se o seu marido/parceiro usava/usa camisinha para o coito anal, antes da penetração vaginal ele a retirava ou trocava/retira ou troca?
¦1¦ Sim ¦2¦ Não ¦3¦ As vezes
¦4¦ Não usava/usa camisinha

105. O seu marido/parceiro tinha/tem o hábito de praticar sexo oral na Sra., ou seja, contato da boca ou lingua dele nos seus genitais?
¦1¦ Sim, frequentemente ¦2¦ Sim, raramente ¦3¦ Não

106. (*Se sim, perguntar*) Com quantos parceiros a Sra. já praticou/pratica sexo oral desta maneira? _____

Se não, vá direto ao final

Se mais de 1 na resposta anterior, iniciar as próximas perguntas enfatizando que a entrevistada deve se referir ao parceiro com quem ela mais frequentemente praticou/pratica sexo oral.

107. (*Se sim, perguntar*) O seu marido/parceiro realizava/realiza penetração vaginal em seguida ao sexo oral?
¦1¦ Sim ¦2¦ Não ¦3¦ As vezes

Eu agradeco muito a sua colaboração com a nossa pesquisa. Se a Sra. tiver alguma pergunta, sinta-se a vontade em fazê-la. Caso queira comunicar-se comigo depois a Sra pode me procurar aqui durante a semana.

Horário de término da entrevista: ____:____

COMENTARIOS DA ENTREVISTADORA:

_____

_____

_____

_____

_____

Enfermeira: _____

Questionário codificado em ____/____/____ por _____

Dados digitados em ____/____/____ por _____

Dados conferidos em ____/____/____ por _____

APPENDIX 2

CONTENT SUMMARY OF THE BASELINE QUESTIONNAIRE

Content summary of the baseline questionnaire

Question Numbers	Information
1-18	age ethnicity marital status job titles last 10 yrs schooling religion income household goods neighbourhoods of residence where subject lived for the longest time
19-30	*tobacco consumption:* frequency, type, duration, cessation
31-37	*alcohol consumption variables:* specifically type of beverage, frequency, duration
38-53	age at menarche last menstrual period type of menstrual absorbent history of gynecologic symptoms and treatments hygiene habits sexually-transmitted diseases previous Pap smears
54-62	age at first intercourse number of pregnancies and outcomes sexual practices during pregnancy and after delivery
63-87	numbers and types of sexual partners frequency of sexual activity during various periods in the subject's past
88-99	*contraceptive methods:* history of oral contraceptive use frequency of condom use other barrier methods
100-107	practice of anal and oral intercourse.