

Design and methods of the Ludwig-McGill longitudinal study of the natural history of human papillomavirus infection and cervical neoplasia in Brazil

Eduardo Franco,¹ Luisa Villa,² Thomas Rohan,³ Alex Ferenczy,⁴ Maria Petzl-Erler,⁵ and Greg Matlashewski,⁶ for the Ludwig-McGill Study Group⁷

ABSTRACT

This article reports on a large longitudinal study, begun in 1993, of the natural history of human papillomavirus (HPV) infection and cervical neoplasia in a population of low-income women in São Paulo, Brazil, a city with one of the highest risks worldwide for cervical cancer. Known as the Ludwig-McGill cohort study, the epidemiological investigation focuses on persistent infection with oncogenic HPV types as the precursor event leading to cervical neoplasia.

The objectives of this study are to: 1) study the epidemiology of persistent cervical HPV infection in asymptomatic women, 2) investigate whether persistent HPV infection increases risk of low-grade and high-grade cervical lesions, 3) search for determinants of persistent HPV infection, 4) search for molecular variants of HPV that may be associated with an increased risk of lesions, 5) investigate whether viral burden is correlated with persistent infections and with lesion risk, 6) study the antibody response to HPV as a predictor of persistence and lesion progression, and 7) examine the role of HLA typing and codon 72 p53 gene polymorphism in mediating HPV persistence and lesion severity.

The study accrued 2 528 female subjects through March 1997. Subjects were followed up every 4 months in the first year, with twice-yearly return visits to take place in subsequent years. Participants undergo a questionnaire-based interview, have a cervical specimen taken for Pap cytology and HPV testing, and have a blood sample drawn for HPV antibody testing. A cervicography is performed once in the first year and every two years thereafter. In this article we describe the design and methods of the study, provide baseline cohort characteristics, and present a preliminary assessment of the prognostic value of baseline HPV status.

¹ McGill University, Department of Oncology, and Department of Epidemiology, Montreal, Quebec, Canada. Address correspondence and reprint requests to: Prof. E. Franco, Department of Oncology, McGill University, 546 Pine Avenue West, Montreal, QC, Canada H2W 1S6. Phone: (514) 398-6032; fax: (514) 398-5002; e-mail: eduardof@oncology.lan.mcgill.ca

² Ludwig Institute for Cancer Research, São Paulo, São Paulo, Brazil.

³ University of Toronto, Department of Public Health Sciences, Toronto, Ontario, Canada.

⁴ McGill University, Department of Pathology, Montreal, Canada, and Sir Mortimer B. Davis Jewish General Hospital, Montreal, Canada.

⁵ Universidade Federal do Paraná, Curitiba, Brazil.

Although few would dispute nowadays that human papillomavirus

⁶ McGill University, Department of Parasitology, Montreal, Canada.

⁷ Additional collaborators: Maria Baggio, Otávia Caballero, João Candeias, Silvaneide Ferreira, Lenice Galan, José Prado, Paula Rahal, Marcella Ribeiro, Antonio Ruiz, Mônica Santos, Laura Sicheiro, João Sobrinho, and Lara Termini (Ludwig Institute); Marie Désy, Eliane Duarte-Franco, Juliette Robitaille, Marie-Claude Rousseau, and Nicolas Schlecht (McGill University).

(HPV) infection is the central cause of cervical cancer, most of the epidemiologic data have come from retrospective, case-control studies (1), which do not provide information on the dynamics of cumulative exposure to cervical HPV infection. Considering the public health and economic importance of cervical cancer and the current widespread interest in the development of HPV vaccines and in using HPV testing to augment existing cytol-

ogy screening programs, there is a clear need for prospective, long-term multidisciplinary studies of the natural history of HPV infection as it progresses to preinvasive cervical lesions.

In 1993, we began a large longitudinal study of the natural history of HPV infection and cervical neoplasia in a population of low-income women in São Paulo, Brazil, a city with one of the highest risks worldwide for cervical cancer. Our study was designed to answer questions that have not yet been addressed in epidemiologic investigations of this neoplastic disease. The Brazilian investigation focuses on persistent infection with oncogenic HPV types as the precursor event leading to cervical neoplasia. The investigation will attempt to understand attributes of the natural history of viral infection that may be instrumental to the design of primary and secondary strategies for preventing cervical cancer. As a unique feature of our cohort study, persistence of HPV infection is monitored by molecular variant analysis (2) and measurement of viral burden (3), which together provide a much finer level of detail than simple HPV testing and may improve the prediction of the likelihood of lesion progression.

The Ludwig-McGill cohort study is designed to further our understanding of the etiopathogenesis of cervical neoplasia by focusing on the following specific objectives:

- measuring the prevalence and incidence of persistent cervical HPV infection in asymptomatic women
- testing the hypothesis that persistent HPV infection increases risk of low-grade and high-grade cervical lesions
- identifying the epidemiologic determinants of persistent cervical HPV infection
- identifying specific molecular variants of oncogenic types of HPV that are associated with increased risk of cervical lesions
- testing the hypothesis that viral burden in the cervix may be correlated with persistent infections and with low- and high-grade lesions

- studying the humoral immune response to HPV as a predictor of persistent cervical HPV infection and of risk of progression of lesion severity
- searching for specific human leukocyte antigen (HLA) alleles or haplotypes associated with HPV persistence and lesion severity
- testing the hypothesis that a specific *p53* gene polymorphism may confer increased resistance against viral persistence and consequently against development of cervical lesions

The purpose of this article is to present a detailed description of the design and methods of the Ludwig-McGill longitudinal study, to describe the characteristics of the cohort at baseline, and to present a preliminary assessment of the prognostic value of baseline HPV status.

MATERIALS AND METHODS

The study described here is a longitudinal study. It involves repeated measurements on the individual participants over time of such risk factors as lifestyle, nutritional and behavioral attributes, and reproductive health and hygiene variables; intermediate endpoints (cervical HPV infection measured via direct DNA detection as well as indirectly, via serologic testing for HPV antibodies); and outcome (preinvasive cervical neoplasia).

Study population and setting

The study population is derived from women attending a comprehensive maternal and child health program catering to low-income families in the city of São Paulo, Brazil. The city has a population of some 12 million and is the capital of the state of São Paulo, the most populous and most industrialized state of Brazil. The clinic setting where subjects were accrued and are being followed up is the Maternidade Escola Dr. Mario de Moraes Altenfelder Silva Municipal Hospital, which is part of a network of primary,

secondary, and tertiary health care institutions maintained by the municipal health department. This clinic is also known by its shorter, unofficial name, "Maternidade Escola Vila Nova Cachoeirinha" (MEVNC).

Eligibility

Women were eligible to participate if they: 1) were between 18 and 60 years old, 2) were permanent residents of the city of São Paulo, 3) were not currently pregnant and had no intention of becoming pregnant during the next 12 months, 4) had an intact uterus and no current referral for hysterectomy, 5) reported no use of vaginal medication in the previous 2 days, and 6) had not had treatment for cervical disease by electrocoagulation, cryotherapy, or conization (the prevailing methods at MEVNC) in the previous 6 months. In addition to these criteria, women were considered ineligible if they were not interested in complying with all scheduled returns, at least for the subsequent 2 years.

Subject recruitment

Two nurses were employed and trained specifically for the study. They recruited subjects by selecting women at random from the daily lists of outpatients in the family medicine, gynecology, and family planning clinics at MEVNC. The nurse-interviewers approached each selected patient to determine eligibility and to explain the general purpose and nature of the study. Women who were potentially eligible were then given a more detailed overview of the study, including the need for cervical and blood specimens, cervicographies, and interviews dealing with sensitive questions, all during multiple scheduled visits to the clinic. The nurses then explained the meal ticket system for compensating participants for the time that they would invest in the study (see the section below on encouraging compliance).

Subjects entered the study only after giving signed informed consent. All

study procedures and the informed consent were approved by the institutional review boards and ethical committees of the participating institutions: McGill University, Montreal, Quebec, Canada; the University of Toronto, Toronto, Ontario, Canada; and the Ludwig Institute for Cancer Research and the MEVNC clinic, both in São Paulo, Brazil.

Scheduled returns and procedures

Subjects enrolled into the study are being followed up over a 5-year period in prescheduled return visits, which will extend into the year 2002. All participants were seen every 4 months in the first year, and are to be seen twice yearly thereafter. In the first four visits and on the annual returns, subjects complete an interviewer-administered structured questionnaire specific for the current visit and have cervical specimens taken for Pap cytology and HPV testing. A 10-cm³ blood sample is also drawn by venipuncture into a vacutainer tube without anticoagulant. During the semester returns between annual follow-up visits, only the cervical specimen is collected for cytology and HPV testing. A cervicography is performed once in the first year during any one of the quadrimester visits. Additional cervicographies are performed at 24 and 48 months.

Encouraging compliance with scheduled follow-up returns

Because of the importance of retaining subjects for the entire duration of the study, participants were told at enrollment that they would receive cash-equivalent incentives. We opted to use meal tickets, which are widely used in Brazil as employee benefits for salaried workers. Meal tickets have an inflation-adjusted cash value and come in various denominations, being honored in nearly all commercial establishments in exchange for meals, groceries, and other goods. The cash-equivalent incentives begin at US\$ 5.00 at the en-

rollment visit and increase \$5 per subsequent visit to a maximum of \$20, which is then given for all subsequent appointments that are kept by the participant. This strategy has ensured excellent rates of follow-up compliance despite the complexity of the procedures used in the study and the requirement for blood specimens.

Questionnaires

In-person interviews are administered by the study nurses according to the schedule described above. Prior to commencement of the study, the nurses underwent extensive training in interview strategies. During these sessions we used role playing, mock interviews, and case studies to train the nurses in how to establish rapport with the respondent, to earn her respect, to avoid being judgmental, and to facilitate recollection of life events, some of which are sensitive. The information that is collected during the interviews covers all classes of risk factors for HPV infection and cervical neoplasia, that is, sociodemographics, reproductive health, sexual practices, smoking, and diet.

Cervical cell specimens

An Accelon biosampler (Medscand, Inc., Hollywood, FL, United States of America) is used to collect a sample of ectocervical and endocervical cells at each of the visits. After the smear is prepared on a glass slide and fixed in 95% ethanol, the sampler containing the exfoliated cells is immersed in a tube containing Tris-EDTA buffer pH 7.4. The tube is agitated to release the cells from the sampler, and the sampler is discarded. Samples are kept at 4 °C at the clinic for 5 days at most and then brought to the laboratory at the Ludwig Institute, where they are kept frozen at -20 °C until testing. The Pap smears are fixed in absolute ethanol, stained, and read at the Ludwig Institute's cytopathology laboratory for an initial diagnosis. They are then

shipped to Montreal, where they are coded and then sent to the laboratory of Dr. Alex Ferenczy, at the Jewish General Hospital, one of McGill University's teaching hospitals. Cytopathology reports produced locally in São Paulo are based on the old Papanicolaou class system, which is the report format preferred by local physicians. The Montreal cytopathology reports are based on the Bethesda system for cytological diagnoses (4).

Cervicography

Because of possible concerns with false negative cytology results, all women also undergo a cervicography during one of the visits in their first year of participation in the study—at a time that is mutually convenient for the participant and for the nurses—and then at 24 and 48 months. Cervicography is being used to detect clinically relevant lesions that are visually identifiable, thus providing a "safety net" to supplement the information obtained from the two cytological readings in the study (see the section below about management of lesions). Cervicography was proposed as a cervical cancer screening tool by Stafil (5) and plays a useful role in large-scale studies in high-risk populations, particularly in remote areas, where well-trained colposcopists cannot be recruited easily. Log sheets and rolls of film are prepared according to the instructions from the provider, National Testing Laboratories (NTL), and then shipped monthly to Fenton, Missouri, United States of America, for development and evaluation by NTL's expert colposcopists. The list of results are then mailed to the project manager in Montreal for computer data entry.

Management of lesions

When women are found to have lesions with moderate dysplasia or worse in the initial reading of their Pap smears in São Paulo, they are referred for colposcopy at MEVNC. Recall is

also triggered if the cytopathology review performed in Montreal reveals a diagnosis of a high-grade squamous intraepithelial lesion (HSIL) or worse. Similarly, cervigrams indicating high-grade lesions or worse are grounds for recall. Since the slide review and the cervicography result are first available in Montreal, we use electronic mail to notify the nurses in São Paulo whenever an HSIL is found. Cumulative lists of HSIL-positive cases are subsequently transferred via a file transfer protocol (FTP) computer connection from Montreal to São Paulo for further verification. This ensures a rapid and safe means of recalling patients with clinically relevant lesions. In the interest of expediency, our own nurses take responsibility for scheduling appointments for the colposcopy, which is performed by the MEVNC gynecologist on duty. At colposcopy, if any lesional tissue is present, a biopsy is taken for histopathological assessment and, if indicated, these women are treated according to the local prevailing protocol. Women with lesions continue to be followed in the cohort if they so wish but do not contribute further person-time to the cohort. All histopathological slides are read locally and then sent to Dr. Ferenczy's laboratory in Montreal for review.

HPV DNA detection

DNA samples are purified by spin column chromatography. Early in the study we extracted DNA from all cervical specimens using digestion with 100 µg/mL proteinase K for 3 h at 55 °C, followed by organic extraction and ethanol precipitation. Cervical specimens are tested for the presence of HPV DNA by a previously described polymerase chain reaction (PCR) protocol amplifying a highly conserved 450 bp segment in the *L1* viral gene (flanked by primers MY09/11) (6, 7). Typing of the amplified products is performed by hybridization with individual oligonucleotide probes specific for all 27 HPV genital types whose nucleotide sequences for probes within

the MY09/11 fragment have been published in the literature. Twenty-four of these have received a taxonomic entry, as HPV 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. Three additional genital types (MM4, MM7, and MM8) are still awaiting taxonomic classification (7). Amplified products that hybridize with the generic probe but with none of the type-specific probes are considered positive for HPV of unknown type(s). These products are tested further, to distinguish among unknown HPVs, by restriction fragment length polymorphism (RFLP) analysis of the *L1* fragment (8). Use of the RFLP analysis extends the range of identifiable HPV types in our study to over 40 genital types, by allowing additional detection of HPVs 32, 34, 44, 62, 64, 67, 69, 70, 72, CP6108, CP8061, CP8304, IS39, and other unknown types. To verify the specificity of the hybridizations, we include more than 30 type-specific positive controls in all membranes. In order to check the integrity of the host DNA material extracted from the specimens, assays also include an additional set of primers (GH20 and PC04) to amplify a 268 bp region of the β -globin gene (6). All HPV assays are done blindly on coded specimens, with no identification linking specimens from the same woman. Appropriate precautions are taken to reduce the possibility of specimen contamination.

Molecular variant analysis

For HPVs 16 and 18, we use a PCR sequencing method with primers flanking a 364 bp segment (nucleotide positions 7478 to 7841) within the LCR region, a hypervariable genomic segment of HPV (9, 10). The amplified products are cloned using the SureClone™ plasmid cloning kit (Pharmacia Biotech, Uppsala, Sweden), according to the manufacturer's instructions and are introduced into *E. coli* XL1-Blue. Recombinant plasmid DNAs are isolated from positive clones and submitted to DNA sequencing by the dideoxy method. Due to possible misincorporation of

bases by *Taq* polymerase and the possibility of multiple coexisting variants, a minimum of five clones are sequenced for each isolate. We also test for molecular variants of HPV 16 in the *E6* gene by targeting a 456 bp segment with nearly the same intratype variability as that in the LCR region that we sequence in our standard protocol. We use the method described by Yamada et al. (11) for sequencing this *E6* fragment. By focusing on two genomic sequences of HPV 16, the most common oncogenic type in our population, we will be able to obtain an extra level of confidence in assigning the putative phylogenetic relatedness for a given variant.

Measuring viral burden

For all cervical specimens found to be positive with the main PCR protocol (MY09/11), we retest them by a quantitative, low-stringency PCR in order to measure viral burden in exfoliated cervical cells (3). The method uses general primers (GP5/6) from a well-known PCR protocol that detects a broad spectrum of HPVs (12). The quantitative PCR protocol employs low-stringency conditions to co-amplify the specific HPV DNA fragment along with DNA sequences from the human genome present in the starting PCR mixture. Standards consisting of mixtures containing varying amounts of reference HPV 16 plasmid (provided by Dr. E. M. de Villiers, Deutsches Krebsforschungszentrum, Heidelberg, Germany) added to a constant background of normal human DNA (corresponding to 0, 4, 20, 100, 500, and 2 500 viral copies per cell) are included in duplicate in every assay. In addition, control samples consisting of DNA from two cervical carcinoma cell lines with known quantities of HPV copies (HeLa, 20–40 copies of HPV 18; Caski, 400–600 copies of HPV 16) are included in duplicate in every assay. The silver-stained gel bands corresponding to the HPV and to the constant human genome fragments are quantified by densitometry (3). The logarithm of the ratio between these two bands is di-

rectly proportional to the logarithm of the amount of HPV DNA in the individual samples. Proper quantification is obtained by linear interpolation in a standard curve constructed with the results from the control mixtures.

Serologic testing for HPV antibodies

Serum samples are separated from the clotted blood specimens and stored at -20°C until testing. An ELISA technique described previously (13) is used for the semiquantitative assessment of immunoglobulin G (IgG) antibodies to L1 and L2 HPV capsid antigens. Antigen preparations consist of self-assembled empty virus-like particles (VLP) produced from a baculovirus system expressing L1 and L2 (14). A sufficiently large batch of HPV 16 VLP antigen was donated by Dr. John Schiller, Laboratory of Cellular Oncology, U.S. National Institutes of Health. This initial batch is being used to define seropositivity at baseline and initial follow-up visits. Subsequent batches are being prepared at the Ludwig Institute using strict adherence to Dr. Schiller's expression system protocol. Seropositivity is arbitrarily based on the cutpoint for the 90th percentile of the distribution of reactivity for control specimens from women free of lesions and testing consistently negative for HPV DNA during all four first-year visits. Advances in VLP technology permitting, we will incorporate in our assay system VLP antigens from any other HPV types that become available.

Testing for host susceptibility markers

One of the most important objectives of research on the prevention of cervical cancer is to identify host cofactors that could influence HPV carcinogenesis. We are testing for two such cofactors, the woman's HLA makeup and *p53* polymorphism in codon 72. HLA class II alleles or haplotypes may play

a role in the HPV infection and genetic susceptibility to cervical cancer, probably regulating the immune response against HPV infection and ultimately interfering in the establishment of productive persistent infections.

One pivotal step in the etiology of HPV-induced cervical neoplasia seems to be the binding of the viral *E6* gene product to the *p53* tumor suppressor protein. This promotes degradation of the latter, exempting the cell to undergo *p53*-mediated control. In the general population, there are two common and distinct wild-type alleles of the *p53* tumor suppressor gene that are not functionally equivalent with respect to HPV. These alleles represent a polymorphism. One encodes arginine at amino acid residue 72 (*Arg* allele), and the other encodes a proline at amino acid residue 72 (*Pro* allele). At the molecular level, HPVs can mediate the degradation of the *p53Arg* protein more efficiently than the *p53Pro* protein (15), raising the possibility that individuals homozygous for the *p53Arg* alleles would be more susceptible to HPV-associated pathology than individuals with *p53Pro* alleles. We hypothesize that women with the homozygous *p53Arg* genotype will be more susceptible to the acquisition of persistent HPV infection and, ultimately, to the development of cervical intraepithelial neoplasia than those homozygous or heterozygous for the *p53Pro* allele. It is possible that in women with the homozygous *p53Arg* genotype, HPV infection will be established more productively with colonization of the cervical epithelium because the virus will be more successful in inhibiting *p53*-mediated apoptosis in infected cells. Hence, there is the expectation that these women will be more susceptible to persistent HPV detection.

The hypothesis regarding the role of a *p53* polymorphism is at present highly controversial. The original association (15) was not confirmed by subsequent studies, such as by Hildesheim et al. (16). However, we have obtained preliminary data from a case-control study conducted in Brazil that

indicates a greater cervical cancer risk among women with the homozygous *p53Arg* form. Nevertheless, we were able to identify the association only after restricting the analysis to subjects with concordant genotype information obtained in three different laboratories. This suggests that misclassification of genotypes could be statistically diluting the association, thus preventing it from being measured in epidemiologic studies (Makni et al., manuscript submitted).

HLA typing. Molecular typing of HLA alleles will be performed in the DNA extracted from the enrollment cervical specimen, or, when that specimen is depleted or insufficient, from any of the cervical specimens collected in subsequent visits. The HLA class II DRB1 and DQB1 high-resolution typing will be performed by PCR and sequence-specific oligonucleotide probe hybridization following the standard international nomenclature (17). All HLA typing will be performed blindly on coded samples.

***P53* polymorphism analysis.** The *p53* genotype for codon 72 will be determined on the enrollment DNA samples (or using subsequent ones if the first specimen is depleted) using a recently described PCR technique (15). In brief, two PCR reactions will be carried out on each sample, one with a primer set specific to the *p53Pro* allele and the other with a primer set specific to the *p53Arg* allele. The *p53Pro* allele yields a 277 bp fragment, and the *p53Arg* allele yields a 141 bp fragment. Primer pairs (2 μg each) are mixed and end labeled with [^{32}P] using polynucleotide kinase. PCR reactions (25 cycles) are performed in a volume of 50 μL using 0.2 units of Red Hot polymerase (Advanced Biotechnologies, Surrey, United Kingdom) in the presence of 1.5 mM Mg^{++} and in a reaction buffer supplied by the manufacturer. An aliquot of the reaction product will be fractionated on a 6% polyacrylamide gel, dried, and exposed to

TABLE 2. Characteristics of subject participation and reasons for ineligibility at time of initial enrollment, São Paulo, Brazil, accrual period 1993–1997

Status	Category	% (in relation to)		
		Number	Status	Overall
Ineligible	Presently pregnant or intends to become	411	33.0	8.2
	Total or partial hysterectomy	285	22.9	5.7
	Residence outside São Paulo or intention to move soon	223	17.9	4.5
	Age < 18 or > 60	158	12.7	3.2
	Reasons invalidating Pap smear	91	7.3	1.8
	Recent surgery preventing participation	40	3.2	0.8
	Virgin	20	1.6	0.4
	Mental retardation or deafness	17	1.4	0.3
	Total	1 245	100	24.9
Unknown	Refused to be approached	138	88.5	2.8
	Did not provide information when approached	18	11.5	0.4
	Total	156	100	3.1
Eligible	Enrolled in the study	2 528	70.4	50.7
	Refused participation	1 061	29.6	21.3
	Total	3 589	100	71.9
Total (all categories)		4 990	NA ^a	100

^a NA = not applicable.

(median: 32; mode: 30 years). Most subjects were white and of European ancestry (Portuguese, Italian, German, Arab, Spanish). More than 80% of the women had not attended high school. However, reflecting the clinic-based sampling used for recruitment, the vast majority of the subjects reported having had Pap tests previously. A little over one-third of all women reported being current smokers (Table 3), which suggests that this study will attain a relatively high statistical power to investigate the role of smoking in the epidemiology of cervical neoplasia, a currently unresolved issue.

Baseline cervical cytology and cervicography results

Enrollment cervical specimens were evaluable for 2 514 women. For the remaining 14 women, the specimen was insufficient to enable appropriate cytological review to be performed, or the slides were lost or broken. Table 4 shows the distribution of subjects with respect to their baseline cytology results, as determined from the review

of the enrollment Pap smears at McGill University. The prevalence of SILs of both low and high grade was 2.2%.

A blind assessment of the diagnostic reproducibility was done in Montreal with random samples of 109 normal and 103 SIL (74 LSIL, 29 HSIL) smears. The kappa statistic was 0.847, with only 19 of 212 smears (9.0%) classified outside of their original diagnostic categories (normal, LSIL, HSIL), and only 10 of 212 (4.7%) if diagnostic mismatches were based on a broader, any-grade SIL category.

The local cytology reading in São Paulo was based on the old Papanicolaou classification system. The percentages by class were as follows: I (normal): 26.7%; II (inflammatory): 71.2%; III (not otherwise specified): 1.3%; IIIa (mild dysplasia): 0.7%; IIIc (severe dysplasia = 1 case): 0.0%; IV (carcinoma in situ): 0.1%; and V (invasive cancer = 1 case): 0.0%. Of the 86 cases of SIL detected either by the local smear reading or by the cytology review in Montreal, only 23 were detected at both opportunities.

Cervicography was introduced in the study only in December 1994,

which prevented 628 women from being examined during the first year. The remaining 1 886 women were classified as follows: 1 616 (85.7%) within normal limits, 214 (11.3%) low-grade lesions A or B, 9 (0.5%) high-grade lesions, and 47 (2.5%) unsatisfactory. Of the 223 low-grade or high-grade lesions detected by cervicography, only 4 (1.8%) were also diagnosed as Pap classes III or higher by the local cytology, and 9 (4.0%) were labeled as LSIL or HSIL by the cytology review.

Baseline HPV status

As of August 1998, we had completed HPV DNA testing and typing for 4 935 specimens from 1 430 women, representing the first-year specimens from those accruing long-term follow-up. The prevalence of HPV infection by any type at entry was 13.8%. The most prevalent types were HPV 16 (2.8%), HPV 53 (1.5%), HPV 58 (1.2%), HPV 6/11 (1.0%), HPV 31 (1.0%), HPV 70 (0.9%), and HPV 18 (0.8%). Among HPV-positive specimens, the mean viral copy number was 3.0 per cell, with most specimens having 1 or fewer viral copies per cell. Only 11% of the HPV-positive specimens at enrollment had more than 100 copies per cell.

Determination of molecular variants of HPV by sequencing is the most time-consuming procedure in the study. We have thus far found seven different variants of HPV 16 in 97 isolates from 65 subjects. These variants are grouped phylogenetically as follows: E branch: prototype, B-12; AA branch: B-2, B-2A, one new (related to B2); Af2 branch: S21-A; and As branch: AC-2. Four variants of HPV 18 were identified in 27 isolates from 11 subjects: B18-2, B18-6, one new (related to B18-2), and T18-17.

Preliminary assessment of the prognostic value of baseline HPV status

We analyzed the risk of any-grade, incident SIL in the cohort as a function of HPV positivity at enrollment. Data

TABLE 3. Distribution of selected characteristics at enrollment for participants in the Ludwig-McGill cohort, São Paulo, Brazil, accrual period 1993–1997

Variable	Categories	Number ^a	%
Age	18–24	485	20.3
	25–34	944	39.6
	35–44	686	28.8
	≥ 45	270	11.3
Ethnicity	White	1 535	64.4
	Mestiza	511	21.4
	Black	301	12.6
	Other	36	1.5
Education	Less than elementary	543	22.8
	Elementary	1 393	58.5
	High school	381	16.0
	College/university	66	2.8
Smoking	Never	1 137	47.7
	Current	834	35.0
	Former	414	17.4
Age at first intercourse	≤ 15	652	27.3
	16–17	610	25.6
	18–19	501	21.0
	≥ 20	622	26.1
Lifetime number of sexual partners	1	1 058	44.4
	2	502	21.1
	3	331	13.9
	4	192	8.1
	5	115	4.8
	≥ 6	186	7.8
Duration of oral contraceptive use	Never	387	16.2
	≤ 5 years	1 308	54.8
	> 5 years	690	28.9
Number of pregnancies	0	45	1.9
	1	353	14.9
	2	505	21.3
	3	500	21.1
	4–5	568	24.0
	≥ 6	397	16.8
Previous number of Pap tests	0	124	5.2
	1–2	534	22.4
	3–4	509	21.3
	5–6	315	13.2
	≥ 7	903	37.9

^a Information is for 2 385 subjects with enrollment questionnaires processed as of June 1998; missing values are excluded.

on cytological outcomes were based on smears read through August 1997. Only the 887 women who were free of SIL at enrollment and had been tested for HPV were included in the analysis. Two separate incidence curves were computed using actuarial techniques, one for cytological outcome deter-

mined by the smear reading by the local hospital provider in São Paulo (Figure 1, top) and the other by the cytology review in Montreal (Figure 1, bottom). The actuarial curves of SIL incidence based on the local cytology were largely overlapping, but not the ones based on the Montreal slide re-

view. Using the research-quality cytology (smear readings provided by Dr. Ferenczy's Montreal laboratory), the relative risk (RR) of SIL associated with HPV positivity was 5.8 (95% confidence interval [CI]: 3.0–11.1) (using the Cox proportional hazards model). Based on the local reading of the same smears, the equivalent RR of SIL was a considerably lower 1.4, with a 95% CI of 0.6–3.3.

DISCUSSION

Although there is currently great enthusiasm concerning the possible application of HPV testing as an adjunct to Pap cytology screening for cervical cancer, there are several problems that need to be solved before any current secondary prevention programs can be augmented. Positive predictive values of HPV testing are low in most asymptomatic women because of the relatively high prevalence of subclinical HPV infection in the general population. Most of these infections are transient and are probably of little significance. The concern resides, however, with the small proportion of women who harbor persistent HPV infections. These women have a much greater risk of subsequent cervical neoplasia (18, 19), indicating that persistent HPV infections, rather than transient ones, are the actual biologic precursor in cervical carcinogenesis. Research on the epidemiology of viral persistence and on its determinants will help formulate algorithms and policies for inclusion of some form of HPV testing as an adjunct to screening for cervical cancer prevention. More importantly, however, we believe that the present study will provide new insights into the pathogenic mechanisms of HPV infection leading to cervical cancer, which will eventually aid in efforts at vaccine development and primary prevention.

As far as we know, the Ludwig-McGill longitudinal study is one of the largest ongoing epidemiologic investigations of the natural history of cervical neoplasia to employ a powerful array of laboratory techniques to doc-

TABLE 4. Distribution of cervical cytological results (based on slide review) at enrollment among cohort participants, São Paulo, Brazil, accrual period 1993–1997

Cytological diagnosis ^a	Number	Percent	Total (%) ^b
Within normal limits:			2 411 (96.1)
Normal	1 627	64.7	
Benign cellular changes	784	31.2	
ASCUS:			43 (1.7)
ASCUS, favor benign cellular changes	8	0.3	
ASCUS, rule out LSIL	29	1.2	
ASCUS, rule out HSIL	5	0.2	
AGUS, rule out adenocarcinoma	1	0.0	
LSIL:			33 (1.3)
LSIL	30	1.2	
LSIL, rule out HSIL	3	0.1	
HSIL:			21 (0.8)
HSIL, moderate dysplasia	7	0.3	
HSIL, severe dysplasia	14	0.6	
Unsatisfactory	6	0.2	

^a Abbreviations: ASCUS = atypical squamous cells of undetermined significance; AGUS = atypical glandular cells of undetermined significance; LSIL = low grade squamous intraepithelial lesions; HSIL = high grade squamous intraepithelial lesions.

^b Percent totals by diagnostic grouping exclude unsatisfactory smears from the denominator. Overall and diagnostic group percent totals do not necessarily add up because of rounding errors.

ument the onset of precursor lesions. It is also one of the few ongoing cohort studies with multiple measurements of HPV markers over time, thus allowing determination of the true incidence rate of viral infection. Since it is being conducted in a population at high risk for cervical cancer, the levels of statistical power for cytological endpoints might be higher than if the same study were conducted in a typical North American population.

Subject recruitment was completed in March 1997, after the cohort reached the sample size that we had estimated was needed to attain sufficient statistical power to assess associations of interest. Actuarial rates of compliance with all follow-up visits have stabilized at around 75%. Such a rate of long-term adherence can be considered as very high in a protocol with procedure-intensive follow-up that requires multiple hospital visits frequently lasting longer than one hour. This compliance rate has been attained despite the fact that participation is dependent upon agreeing to donate a blood sample and a cervical cell specimen and to

undergo lengthy interviews dealing with sensitive life events in most clinic visits.

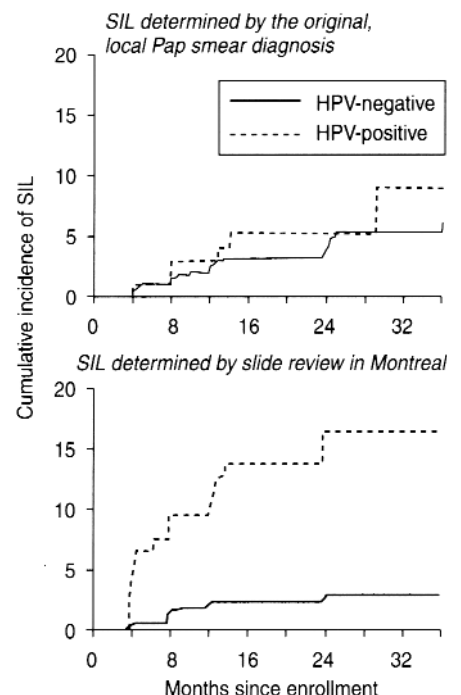
An ongoing cohort study in Costa Rica also includes repeated testing of subjects for HPV and cervical neoplasia over time, but the focus of that study is on the role of HPV in the etiology of HSIL and on the evaluation of new cervical cancer screening strategies (20). Unlike other cohort studies that use viral type as a taxonomic unit, we have chosen the innovative approach of testing for molecular variants and measuring viral load in the cervix to study viral persistence. We will further document the natural history of cervical HPV infection as the precursor event for SIL by correlating the viral findings with the subject's immune response to specific capsid antigens of HPV and by correlating the latter with risk of subsequent SIL.

The focus of our investigation is on the dynamics of the natural history of HPV infection and cervical neoplasia. Although it is not one of our main objectives, our cohort study will provide a limited assessment of the relative diag-

nostic performance of Pap cytology, cervicography, and HPV testing, as well as their combinations, using the biopsy results as the gold standard. We will calculate the sensitivity and specificity for each combination of test and lesion severity. Women who were recalled as a result of the algorithm process described earlier in the "Management of lesions" subsection but who did not have lesional tissue to be biopsied will be considered disease-free for the purposes of the latter calculation.

Our preliminary assessment of cytologic outcomes in the study indicates that there was substantial underestimation of the prognostic value of HPV status at entry with respect to lesion outcome using the nonresearch-quality smear reading, presumably as a reflection of outcome misclassification. This empirical illustration of the effects of cytological misclassification prompted us to stop using the cytology diagnoses from the local hospital provider for

FIGURE 1. Cumulative incidence of any-grade squamous intraepithelial lesions (SIL) associated with baseline HPV positivity among 887 women with no cytological abnormalities when enrolled in the Ludwig-McGill cohort study, São Paulo, Brazil, 1993 and after



triage purposes. An accredited cytotechnologist has joined our team to provide a timely and reliable first screen of all smears generated in the study, to prevent delays in referring for colposcopy any cases of HSIL.

The traditional epidemiologic study designs of single-opportunity assessment of exposure and outcome do not permit assessment of questions of viral persistence, fluctuation in viral load, regression of cervical lesions, and the dynamics of risk factor changes over time (e.g., acquisition of

new sexual partners) (21). To understand the role and mechanism of such dynamic changes in the natural history of the disease one must conduct studies that repeatedly collect data on risk factors, HPV, and cervical lesions on multiple occasions during follow-up. A longitudinal, repeated-measurement cohort investigation, such as the one used in the Ludwig-McGill study, is the only design that permits accurate and unbiased assessment of cumulative HPV exposure and lesion outcome history.

Acknowledgments. This project is supported by an intramural grant from the Ludwig Institute for Cancer Research and by grants from the U.S. National Cancer Institute, the Medical Research Council of Canada, and the National Cancer Institute of Canada. The authors are indebted to Drs. Maria Nozaki and Lúcia Aoki for the local cytology readings during the initial phase of the investigation. The authors are grateful for the kind contributions of materials by Dr. E. M. de Villiers and by Dr. John Schiller.

REFERENCES

- IARC Working Group. Human papillomaviruses. Lyon, France: International Agency for Research on Cancer; 1995. (IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Vol. 64).
- Franco EL, Villa LL, Rahal P, Ruiz A. Molecular variant analysis as an epidemiological tool to study persistence of cervical human papillomavirus infection. *J Natl Cancer Inst* 1994; 86(20):1558-1559.
- Caballero OL, Villa LL, Simpson AJG. Low stringency-PCR (LS-PCR) allows entirely internally standardized DNA quantitation. *Nucleic Acid Res* 1995;23:192-193.
- 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.
- Stafil A. Cervicography: a new method for cervical cancer detection. *Am J Obstet Gynecol* 1981;139:815-825.
- Bauer HM, Ting Y, Greer CE, Chambers JC, Tashiro CJ, Chimera J, et al. Genital human papillomavirus infection in female university students as determined by a PCR-based method. *JAMA* 1991;265:472-477.
- Hildesheim A, Schiffman MH, Gravitt PE, Glass AG, Greer CE, Zhang T, et al. Persistence of type-specific human papillomavirus infection among cytologically normal women. *J Infect Dis* 1994;169:235-240.
- Bernard HU, Chan SY, Manos MM, Ong CK, Villa LL, Delius H, et al. Identification and assessment of known and novel human papillomaviruses by polymerase chain reaction amplification, restriction fragment length polymorphisms, nucleotide sequence, and phylogenetic algorithms. *J Infect Dis* 1994;170:1077-1085.
- Ho L, Chan SY, Chow V, Chong T, Tay SK, Villa LL, et al. Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree. *J Clin Microbiol* 1991;29:1765-1772.
- Ong CK, Chan SY, Campo MS, Fujinaga K, Mavromaranazos P, Labropoulou V, et al. Evolution of human papillomavirus type-18: an ancient phylogenetic root in Africa and intratype diversity reflect coevolution with human ethnic groups. *J Virol* 1993;67: 6424-6431.
- Yamada T, Wheeler CM, Halpern AL, Stewart ACM, Hildesheim A, Jenison SA. Human papillomavirus type 16 variant lineages in United States populations characterized by nucleotide sequence analysis of the E6, L2, and L1 coding segments. *J Virol* 1995;69:7743-7753.
- VanDenBrule AJC, Snijders PJF, Gordijn RLJ, Bleker OP, Meijer CJLM, Walboomers JMM. 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.
- Kirnbauer R, Hubbert NL, Wheeler CM, Becker TM, Lowy DR, Schiller JT. A virus-like particle enzyme-linked immunosorbent assay detects serum antibodies in a majority of women infected with human papillomavirus type 16. *J Natl Cancer Inst* 1994;86:494-499.
- Kirnbauer R, Taub J, Greenstone H, Roden R, Durst M, Gissmann L, et al. Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 into virus-like particles. *J Virol* 1993;67:6929-6936.
- Storey A, Thomas M, Kalita A, Harwood C, Gardiol D, Mantovani F, et al. Role of a p53 polymorphism in the development of human papillomavirus-associated cancer. *Nature* 1998;393(6682):229-234.
- Hildesheim A, Schiffman M, Brinton L, Fraumeni Jr J, Herrero R, Bratti MC, et al. p53 polymorphism and risk of cervical cancer. *Nature* 1998;396(6711):531-532.
- Bodmer JG, Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Charron D, et al. Nomenclature for factors of the HLA system, 1996. *Vox Sanguinis* 1997;73(2):105-130.
- Koutsky LA, Holmes KK, Critchlow CW, Stevens CE, Paavonen J, Beckmann AM, et al. 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.
- Ho GYF, Burk RD, Klein S, Kadish AS, Chang CJ, Palan P, et al. Persistent genital human papillomavirus infection as a risk factor for persistent cervical dysplasia. *J Natl Cancer Inst* 1995;87:1365-1371.
- Herrero R, Schiffman MH, Bratti C, Hildesheim A, Balmaceda I, Sherman ME, et al. Design and methods of a population-based natural history study of cervical neoplasia in a rural province of Costa Rica: the Guanacaste Project. *Rev Panam Salud Publica* 1997;1(5): 362-375.
- Franco EL, Rohan T, Villa L. Epidemiologic evidence and human papillomavirus infection as a necessary cause of cervical cancer. *J Natl Cancer Inst* 1999;91:506-511.

Manuscript received on 11 January 1999. Revised version accepted for publication on 11 June 1999.

RESUMEN**Diseño y métodos del estudio longitudinal Ludwig-McGill sobre la historia natural de la infección por papilomavirus humanos y la neoplasia cervical en Brasil**

Este artículo describe un amplio estudio longitudinal, iniciado en 1993, acerca de la historia natural de la infección por papilomavirus humanos (PVH) y las neoplasias cervicales en una población de bajos recursos económicos de São Paulo, Brasil, una de las ciudades con mayor riesgo de cáncer cervical en todo el mundo. Esta investigación epidemiológica, conocida como el estudio de cohorte Ludwig-McGill, se centra en la infección persistente por tipos oncogénicos de PVH como el acontecimiento precursor que conduce a la neoplasia cervical. Sus objetivos consistieron en: 1) estudiar la epidemiología de la infección cervical persistente por PVH en mujeres asintomáticas; 2) investigar si la infección persistente por PVH incrementa el riesgo de lesiones intraepiteliales escamosas cervicales de bajo y alto grado; 3) identificar los determinantes de la infección persistente por PVH; 4) buscar variantes moleculares de los PVH que puedan estar asociadas a un aumento del riesgo de lesiones; 5) determinar si la carga vírica está correlacionada con la infección persistente y el riesgo de lesiones; 6) averiguar si la respuesta de anticuerpos frente a los PVH permite predecir la persistencia de la infección y la progresión de las lesiones, y 7) investigar el papel de los tipos HLA y del polimorfismo del codón 72 del gen *p53* como mediadores de la persistencia de los PVH y de la gravedad de las lesiones. Hasta marzo de 1997 se reclutaron 2 528 mujeres que fueron examinadas cada cuatro meses en el primer año y cada seis meses en los años subsecuentes. Las participantes se someten a una entrevista basada en un cuestionario; a la obtención de una muestra cervical para citología de Papanicolaou y detección de PVH, y a la extracción de una muestra de sangre para identificar anticuerpos anti-PVH. Además, en el primer año se les realiza una cervicografía, que después se repite cada dos años. En este artículo se describen el diseño y los métodos del estudio, se presentan las características basales de la cohorte y se realiza una evaluación preliminar del valor pronóstico del estado basal de los PVH.
