RANDOMIZED CONTROLLED TRIAL OF HUMAN PAPILLOMAVIRUS TESTING VERSUS PAP CYTOLOGY FOR PRIMARY SCREENING OF CERVICAL CANCER PRECURSORS

Marie-Hélène Mayrand

Department of Epidemiology and Biostatistics

McGill University, Montreal

August 2007

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Philosophiae Doctor (Ph.D.) in Epidemiology and Biostatistics

Your file Votre référence
ISBN: 978-0-494-50956-2
Our file Notre référence
ISBN: 978-0-494-50956-2

ABSTRACT

Introduction: Despite its successful history in cervical cancer screening, the false negative rate of Papanicolaou (Pap) cytology is an important concern. Testing for DNA of oncogenic human papillomaviruses (HPV) could circumvent this limitation. The present thesis reports the results of the first screening round of the Canadian Cervical Cancer Screening Trial (CCCaST).

Objectives: To compare the performance (sensitivity, specificity, positive and negative predictive values) of HPV testing vs. Pap cytology in the detection of cervical cancers and their high-grade precursors and explore the impact of sampling order, test thresholds and combinations, and patient characteristics and laboratory on test performance.

Methods: CCCaST is a randomized controlled trial (RCT) that compares HPV testing with Hybrid Capture 2 and conventional Pap cytology as screening interventions to identify high-grade cervical intraepithelial neoplasia among women aged 30-69 years in Montreal, Quebec, and St. John's, Newfoundland. Women with an abnormal Pap (≥ atypical squamous cells) or a positive HPV (≥ 1 pg/ml of high-risk HPV DNA) underwent colposcopy and biopsy, as did a random sample of women with negative tests. Performance estimates were corrected for verification bias. Logistic regression modelling was used to investigate the importance of spectrum effect due to selected patient and laboratory characteristics.

Results: 10,154 women were randomized. The sensitivity of HPV testing was substantially higher (94.6%, 95% confidence interval [CI]: 84.2-100) than that of Pap cytology (55.4%, 95%CI: 33.6-77.2), albeit with a slightly lower specificity (Pap: 96.8%, 95%CI: 96.3-97.3; HPV: (94.1%, 95%CI: 93.4-94.8). Performance was unaffected by which test was performed first. Use of both tests in parallel was 100% sensitive and 92.5% specific. Triage algorithms for Pap or HPV resulted in fewer referrals but were less sensitive. Spectrum effect analysis revealed that women over 40 reporting no new sexual partner in the last year would benefit from the higher sensitivity of HPV testing, while benefiting from comparable specificity to that of Pap cytology.

Conclusion: The first RCT of HPV versus Pap testing for primary screening in North America demonstrates the greater sensitivity of HPV compared to cytology. Our study supports a paradigm change in cervical cancer screening.

RÉSUMÉ

Introduction: Malgré le succès de la cytologie cervicale (test Pap) dans la prévention du cancer du col, les résultats faussement négatifs demeure une limitation importante. La détection d'ADN de virus du papillome humains oncogènes (test VPH) offre la possibilité de surmonter cette barrière. Le présent document résume les résultats de la première phase de l'Étude Canadienne sur le Dépistage du Cancer du Col (CCCaST).

Objectifs: Comparer la performance (sensibilité, spécificité, valeurs prédictives positives et négatives) du test VPH à celle du test Pap pour la détection des précurseurs du cancer du col, et explorer l'impact de l'ordre de prélèvement, des valeurs seuils et combinaisons, et de certaines caractéristiques des participantes sur la performance de ces tests.

Méthodes: CCCaST est un essai randomisé contrôlé (ERC) comparant 2 interventions de dépistage: un test VPH (Hybrid Capture 2) et le Pap conventionnel chez les femmes de 30 à 69 ans à Montréal, Québec, et St. John's, Terre-Neuve. Les participantes ayant un Pap anormal (≥ atypies des cellules épithéliales) ou un test VPH positif (≥ 1 pg/ml d'ADN de VPH à haut risque) ont subi colposcopie et biopsies, tout comme un échantillon aléatoire de participantes ayant des résultats négatifs. Les estimés de performance ont été corrigés pour le biais de vérification. La régression logistique a permis d'explorer l'importance de différentes variables sur la performance des tests.

Résultats: 10,154 femmes ont été randomisées. La sensibilité du test VPH était beaucoup plus élevée (94.6%, intervalle de confiance à 95% [CI]: 84.2-100) que celle du Pap (55.4%, 95%CI: 33.6-77.2), mais sa spécificité légèrement plus faible (Pap: 96.8%, 95%CI: 96.3-97.3; HPV: (94.1%, 95%CI: 93.4-94.8). L'ordre de prélèvement n'a pas influencé la performance des tests. L'utilisation simultanée des 2 tests a présenté une sensibilité de 100% et une spécificité de 92.5%. Les algorithmes incorporant un tri ont été moins sensibles mais ont entraîné moins de références. Les femmes de 40 ans n'ayant pas de nouveau partenaire sexuel dans la dernière année pourraient bénéficier de la sensibilité accrue du test VPH tout en n'ayant pas à subir les inconvénients d'une baisse de spécificité.

Conclusion: Le premier ECR comparant le test VPH au test Pap démontre la meilleure sensibilité du test VPH et ainsi supporte un changement de paradigme pour le dépistage du cancer du col utérin.

PREFACE

This thesis is comprised of an introduction (which includes a review of the topic, the rational for this project and its objectives), a methods section, a detailed presentation of results, followed by a discussion and a conclusion. This thesis follows the traditional format. Although all the material included in the document was written specifically for this thesis, certain sections of the thesis have been summarized and submitted as manuscripts for publication.

Certain sections of chapter 1 (Introduction) were used to prepare a book chapter focusing on uterine cancer prevention. [Mayrand MH, Franco EL. Prevention of Uterine Cancers. In: Meisels A, Morin C eds. Cytopathology of the Uterus, 3r^{d} edition, Chicago, USA. ASCP press, 2007: 136-154.] This book chapter was co-authored by myself and my PhD supervisor, Dr Eduardo Franco. I planned the chapter and the elements to include in it, and wrote the first version. Dr Franco contributed to subsequent and final versions.

A very concise summary of chapter 3 and sections 4.1 and 4.2 (methodology and baseline participants' characteristics) was prepared for publication and accepted [Mayrand MH, Duarte-Franco E, Coutlée F, Rodrigues I, Walter SD, Ratnam S, Franco EL, for the CCCaST Study Group. Randomized study of Human papillomavirus testing versus Pap cytology in the primary screening for cervical cancer precursors: Design, methods and preliminary accrual results of The Canadian Cervical Cancer Screening Trial (CCCaST). International Journal of Cancer 2006; 119:615-23.] This manuscript was co-authored by the co-investigators of CCCaST. I collaborated in the design of the study and assisted with drafting the grant submission that led to funding, under the supervision of my PhD

supervisor, Dr Franco, the principal investigator. I coordinated patient recruitment and data collection in Montreal. I analyzed the data and prepared the first version of this manuscript. E Duarte-Franco, I Rodrigues and S Ratnam collaborated in the implementation of trial procedures. S Ratnam also coordinated patient recruitment and data collection and entry, and HPV testing services in St. John's, Newfoundland. F Coutlée provided infectious disease expertise as well as HPV testing services in Montreal. SD Walter provided statistical expertise for the design and analysis phases. All co-authors provided important intellectual content to revisions of the first draft the manuscript.

Highlights of sections 4.4 and 4.5 were summarized in a very succinct manner and submitted for publication [Mayrand MH, Duarte-Franco E, Rodrigues I, Walter SD, Hanley J, Ferenczy A, Ratnam S, Coutlée F, Franco EL, for the CCCaST Study Group. Randomized controlled trial of human papillomavirus testing versus Pap cytology for primary screening of cervical cancer precursors: results from the Canadian Cervical Cancer Screening Trial (CCCaST)]. I devised the analysis strategy under the supervision of Dr Franco. I performed the statistical analysis and prepared the first version of this manuscript. SD Walter and J Hanley provided statistical expertise. A Ferenczy provided expertise on histological examinations. All co-authors provided important intellectual content to revisions of the first draft the manuscript. This manuscript has been accepted for publication in the New England Journal of Medicine (June 2007).

STATEMENT OF ORIGINALITY

This thesis summarizes the main findings from the first phase of the Canadian Cervical Cancer Screening Trial (CCCaST). CCCaST is the first randomized controlled trial to compare HPV testing to Pap cytology as standalone tests for primary screening of cervical cancer and its precursors.

I have had the chance to be involved in CCCaST from its beginning. Indeed, I collaborated in the elaboration of the protocol, in grant funding submissions, in implementing study procedures, patient recruitment and data collection. I was also the coordinator for the Montreal component of CCCaST.

Although other studies had compared Pap to HPV testing for cervical cancer screening, design limitations and population specificities precluded reaching firm conclusions. CCCaST possesses several unique features which lend strength to its conclusions: it is a randomized trial, it targets a population of women and health care practitioners in the context of routine cervical cancer screening activities, screening tests were performed in regular community settings, the gold standard was applied blindly of screening test results, the design allowed for correction of verification bias.

The main results clearly demonstrate the superior sensitivity of HPV testing for cervical cancer precursor screening. The exploratory analyses provide insights into possible performance of test combination. In the thesis, is also the first report of spectrum effect analysis of Pap and HPV testing, which identified a subgroup of women most likely to benefit from HPV testing.

Given the innovative nature if this trial and the importance of cervical cancer screening, the results summarized in this thesis will no doubt impact public health decision making.

ACKNOWLEDGMENTS

This thesis represents the fruit of 7 years of professional and personal commitment. It would not have been possible without the continued support of a multitalented team. Every person mentioned in this section has contributed to my learning and research experience. Dr Eduardo Franco, as my thesis supervisor and the principal investigator for the CCCaST study, no doubt deserves the first mention. From the start, he believed in my abilities to contribute to this project. Through on-going mentoring, he has helped me develop the skills I will need to succeed as a researcher. The CCCaST co-investigators, Eliane Duarte-Franco, Isabel Rodrigues, Stephen Walter, Sam Ratnam, and François Coutlée, have all in their own ways enriched my doctoral experience. My doctoral thesis committee members, Jean-Paul Collet and Robert Platt, have been available throughout my stay at McGill to assist me, and have provided insightful comments to the different versions of this thesis.

Below is a list of research staff and clinical collaborators who have contributed to the success of CCCaST. Even though I did not have the chance to meet all of them, I sincerely thank them for their participation. Special thoughts to the Montreal CCCaST ladies who have dealt with deadlines, changing work needs, competing priorities, with grace, humour and professionalism.

Montreal Research Staff: L Abruzzese, K Bellegarde, M Bernardin, F Ferdinand, G Kelsal, M Paquin, S Piché, J Poirier, A Rodrigues, N Rousseau, C Schwartz, N Slavtcheva, E Tunistky.

St. John's Research Staff: G Condon, D Mason, B Halfyard, J McGrath, A Batstone, W Shea, C Leonard, E Oates, V Moulton, P Francis, C Head, A Fitzpatrick.

Montreal Clinical Collaborators: MY Arsenault, G Asselin, L Authier, S Bagga, P Bastien, L Bazinet, F Beaudoin, P Beaulieu, MJ Bédard, S Belinski, S Bélisle, J Benoit, M Bernard, S Bianki, L Biron, F Bissonnette, R Bou-Habib, J Bourque, B Bradbury, M Champagne, Y Charles, P Choquette, J N Couture, HQ Dao, J Desjardins, C Desjardins, L Desrosiers, S DiTommaso, L Dontigny, M Doyle, J Dubé, MJ Dupuis, F Durocher, F Engel, B Fafard, A Ferenczy, G Fortier, A Fortin, C Fortin, D Francoeur, D Frechette, P Fugère, G Gagné, GP Gagné, S Gascon, MJ Gaudreau, D Gaudron, K Gemayel, L Gilbert, S Gilbert, J Gill, I Girard, A Gobeil, L Granger, F Grou, E Grou, G Guertin, J Guimond, R Hemmings, N Ifergan, L Johnson, C Johnson, L Ketchian, Y Korcaz, C Lafortune, J Lalande, JF Lanctot, D Landry, M Landry, D Langevin, I Langlois, L Lanmy-Monnot, L Lapensée, L Larouche, D Laurin, MC Lavigne, Y Lavoie, M Leduc, F Leger, N Leroux, G Luskey, N Mansour, J Marceau, A Masse, MH Mayrand, I Mayrand, LR McLauchlin, S Menard, C Mercer, M Messier, B Michon, M Nadeau, M Nguyen, S Ouellet, C Paquin, R Paré, S Peloquin, Y Piché, R Pichet, C Rivard, I Rodrigues, S Roman, L Rusimovic, G Sanche, D Souliere, D Sproule, M Steben, S Still, D Theriault, G Tondreau, D Tremblay, T Minh Dung Vo, VM Whitehead, M Yaffe, A Di Zazzo, C Ziegler.

Newfoundland Clinical Collaborators: L Dawson, D Fontaine, J Dunne, N Kum, C Pike, P Power, P Wadden, E Bannister, E Callahan, J Collingwood, P Crocker, A Drover, F Fifield, J Fitzgerald, B Grandy, M Greene, K Halley, L Hatcher, J Hickey, P Horwood, J

Janes, F Jardine, L Kieley, S King, C Kirby, E Mate, S McGrath, C McManamon, K Misik, M O'Dea, P O'Shea, C Peddle, M Penton, L Rogers, K Saunders, P Skirving, T Sullivan, J Verge, M Watson, M Young.

It is impossible to successfully go through the requirements of a PhD program without the support of fellow students. I thank them for the camaraderie, the laughs, and the snacks that have made the first years of the program so much fun, the study period for the comprehensive examination bearable, and the final years of the PhD program doable. Heartfelt thanks to past and present students and staff members of the Division of Cancer Epidemiology: especially Candida Pizzolongo for unbelievably efficient administrative assistance, Salaheddin Mahmud (assistance with database management and analysis), Keira Robinson (assistance with database management) and Helen Trottier (statistical programming, unshakeable enthusiasm, positive energy…and friendship).

Finally, I wish to thank my family for their continued support. A special thanks to Jeanne, Benjamin and Simone for sharing their mom with such a time consuming project. The most precious thank you, of course, goes to my husband Roch Thibert, who saw us through this challenging period of our lives. I hope the whole family can now enjoy smooth(er) sailing now that I move on to another chapter of my professional life.

STATEMENT OF SUPPORT

This project would not have been possible without the financial contribution of granting agencies. CCCaST was funded by a Canadian Institutes of Health Research (CIHR) grant (MCT-54063) to Eduardo Franco, my thesis supervisor. Also, I had the privilege of receiving a fellowship from the National Cancer Institute of Canada. Infrastructure support to the research activities of CCCaST was also made possible by funding from an endowment from the Cancer Research Society to the Division of Cancer Epidemiology at McGill.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

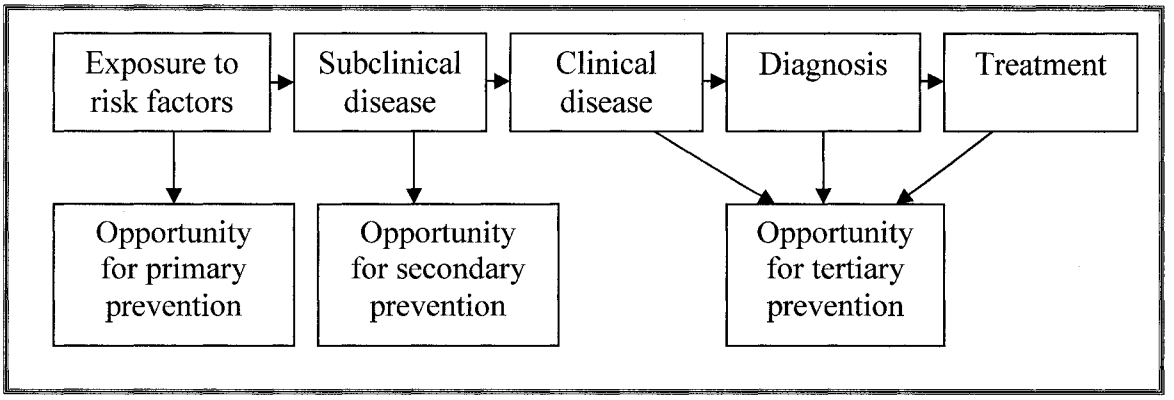
AGC	Abnormal glandular cells
AIC	Akaike's Information Criteria
ASCCP	American Society for Colposcopy and Cervical Pathology
ASC-H	Atypical squamous cells, cannot rule out high grade lesion
AIS	Adenocarcinoma in situ
ASCCP	American Society for Colposcopy and Cervical Pathology
ASC-US	Atypical squamous cells of undetermined significance
CCCaST	Canadian Cervical Cancer Screening Trial
CI	Confidence interval
CIN	Cervical intra-epithelial neoplasia
CIOMS	Council for International Organizations of Medical Sciences
CONSORT	Consolidated standards of reporting trials
DNA	Desoxyribonucleic acid
FDA	Food and Drug Administration
HC2	Hybrid Capture 2
HCT	Hybrid Capture Tube test
HPV	Human papillomavirus
HSIL	High grade squamous intra-epithelial lesion
IARC	International Agency for Research on Cancer
IQR	Inter quartile range
LBC	Liquid-based cytology
LEEP	Loop electro excision procedure
LRM	Logistic regression modelling
LSIL	Low grade squamous intra-epithelial lesion
LR	Likelihood ratio
LRM	Logistic Regression Modeling
NACI	National Advisory Committee on Immunization
NPV	Negative predictive value
PCR	Polymerase chain reaction
PPV	Positive predictive value
RCO	Receiver operator curve
RCT	Randomized controlled trial
RLU	Relative light units
RNA	Ribonucleic acid
STARD	Standards for reporting of diagnostic accuracy
STI	Sexually transmitted infection
STM	Specimen transport medium
VLP	Virus like particules
WMA	World Medical Association

1 INTRODUCTION

1.1 Cancer control: clinical and epidemiological considerations

Strategies to prevent cancer are traditionally divided into three categories: primary prevention, secondary prevention and tertiary prevention (Figure 1). The goal in primary prevention is the reduction of cancer incidence and it can be achieved through the avoidance of the causal factors [Fletcher 1996]. Cancer epidemiology, through the successful identification of risk factors such as smoking, alcohol drinking, diet, and sun light exposure, has contributed to the formulation of many of the current strategies for primary cancer prevention. When the causal agent is infectious, vaccination will also contribute to primary prevention.

Figure 1. Continuum of disease and opportunities for prevention



In secondary prevention, the aim is to detect and treat asymptomatic disease through identification of risk factors or pre-clinical disease. Secondary prevention is most often done through screening, which can be defined as the process whereby one can identify early signs of unrecognized disease (or risk factors) by applying relatively simple clinical

or laboratory tests to individuals at risk of the disease, in order to identify those who should undergo a diagnostic procedure [Hennekens 1987, Fletcher 1996]. Epidemiology plays an important role in assessing the efficiency of secondary prevention programs.

Tertiary prevention strategies are intended to reduce the long-term impact of disease and focus on treatment and follow-up options and their effect on survival, recurrence and quality of life [Fletcher 1996]. Epidemiologic studies, by discovering prognostic factors of unfavourable outcomes, help form the basis of tertiary cancer prevention.

1.1.1 When should we screen

Screening can complement primary prevention or even replace it if primary prevention is infeasible or impractical. However, screening should not be automatically considered appropriate. Indeed, in certain circumstances screening could lead to more harm than good. Over the years a consensus, has developed on the characteristics of diseases, tests and treatments that would warrant screening for a given disease [Miller 1985, Hennekens 1987, Morrison 1992, Fletcher 1996]. They are briefly presented below.

1.1.1.1 Burden of disease:

Since screening entails the use of a significant amount of resources, it is imperative that the disease being screened for represents an important burden in the population, in terms of morbidity and/or mortality.

1.1.1.2 Natural history of disease:

For screening to be worthwhile, the disease should have a reasonably well-understood pre-clinical phase, and the best period for intervention should be easily targeted. In diseases of rapid progression, it will be difficult to identify disease in its pre-clinical phase. On the other hand, diseases with an indolent and protracted pre-clinical phase may never cause problems to the affected individual. In such a case, not only would screening be of no benefit, it could even prove to be detrimental, especially if it led to unnecessary invasive diagnostic and therapeutic measures that may offset in terms of morbidity and mortality risk any potential benefits that may be gained in the long term.

1.1.1.3 Availability of a valid screening test:

The validity of a screening test has several components: (1) its reliability (ability to yield same results if repeated), (2) its accuracy (ability to correctly classify individuals as diseased or not), and (3) its clinical utility (ability to predict the presence or absence of disease in a given individual) [Hennekens 1987, Zhou 2002].

Reliability can be measured by having the same observer repeat its evaluation on the same individual (intra-observer variability) or by having different observers rate the same individual (inter-observer reliability). Measures of agreement, such as Kappa, are often used to quantify such variations. Screening tests that are entirely objective, such as those obtained with a machine in the laboratory, will typically have less variability. To minimize the variability of a screening test which requires human evaluators to reach a conclusion about probable disease status, it is necessary have standardized procures in

place, to train evaluators to follow those procedures and to have regular quality assurance checks [Mahe 2005].

A perfect screening test would classify all diseased individuals as positive and all healthy individuals as negative. Given that a screening test typically produces two possible results: positive and negative, its application in a sample of diseased and non-diseased individuals allows one to construct a 2x2 table with frequencies for all four possible combinations of results: true positives, false positives, false negatives, and true negatives. If the test produces results on a continuous scale, the positive/negative dichotomy can be inferred by applying a proper cut-off value that maximizes the net combination of sensitivity and specificity. For such tests, plotting the sensitivity against "1-specificity" will produce what is called a "receiver-operator curve" (ROC) which may assist in identifying the most appropriate cut-off. Moreover, the area under such curves gives an overall assessment of test performance [Altman 1994b, Zhou 2002].

Table 1 shows the layout and notation for such a table with the formulae for the various indices to assess internal validity. The assessment of the accuracy of a screening test is made most frequently by measuring two primary parameters of test performance: sensitivity and specificity [Taube 1986, Begg 1987, Begg 1991, Zhou 2002]. The sensitivity of a screening test can be defined as its ability to correctly identify diseased individuals. It is expressed as a proportion and measured by dividing the number of true positive tests by the number of diseased subjects. The specificity of a screening test can be defined as its ability to correctly identify healthy (disease-free) individuals. It is expressed as a proportion and measured by dividing the number of true negative tests by

the number of non-diseased individuals. Sensitivity and specificity of screening tests that do not require human interpretation will not vary based on the prevalence of disease. However, screening tests that are subjective may be influenced by prevalence. For example, a screening test which require substantial expertise may have a lower sensitivity in settings with very low prevalence, where evaluators would seldom encounter abnormalities and as such may loose some of there ability to recognize them.

Other parameters, such as Youden's index and the likelihood ratio (LR), gauge the performance of a screening test with respect to its combined ability to yield correct results, positive and negative. Youden's index can yield values between 0 and 1. A value of 0 means that the test gives the same proportion of positive results in diseased and non diseased individuals, and as such has no diagnostic value. A value of 1 means that there are no false positive nor false negative results [Youden 1950, Armitage 1994]. This index is rarely used, mainly because its value is difficult to use to influence health care decisions [Zhou 2002].

Each result of a screening test (i.e. positive and negative in the case of dichotomous tests) has its LR. The LRs is the ratio of the probability of a given test result in people who are diseased to the probability of the same test result in people who are not diseased. Values greater than 1 indicate that the specific test result is associated with disease. Values less than 1 indicate that the specific test result is in favour of an absence of disease. An advantage of the LR is that continuous test results not have to be dichotomized [Zhou 2002, Deeks 2004].

Table 1. Table layout for the calculation of screening performance indices and relevant formulae.

Test result	Disease status	
	Present (D+)	Absent (D-)
Positive (T+)	True positive (TP)	False positive (FP)
Negative (T-)	False negative (FN)	True negative (TN)

Performance index	Formula
Sensitivity (%)	TP / (TP + FN) x 100
Specificity (%)	TN / (TN + FP) x 100
Youden's index	(TP*TN)-(FN*FP)/(TP+FN) (FP+TN)
Positive likelihood ratio*	(TP / (TP + FN))/ FN / (TN + FP)
Negative likelihood ratio*	(FP / (TP + FN))/ TN / (TN + FP)
Positive predictive value (%)	TP / (TP + FP) x 100
Negative predictive value (%)	TN / (TN + FN) x 100

*Formula for dichotomous test

Although sensitivity, specificity and likelihood ratios provide valuable assessments of test performance, they do not help to clarify the health status of a specific patient with a positive or negative test result, and so have limited clinical utility [Altman 1994a]. Predictive values will provide estimates of disease probability given a specific test result, and as such will be a more useful guide for clinical management. The positive predictive value (PPV) can be defined as the probability that someone with a positive test result is in fact diseased. It is expressed as a proportion and can be obtained by dividing the number of true positive tests by the total number of positive tests. The negative predictive value

(NPV) can be defined as the probability that someone with a negative test result is in fact healthy. It is expressed as a proportion and can be obtained by dividing the number of true negative tests by the total number of negative tests. Unlike sensitivity and specificity, the positive and negative predictive values are directly affected by the prevalence of the disease and the values reported in one study do not apply to all settings [Altman 1994a, Armitage 1994]. PPV is mostly influenced by the specificity of the test and the prevalence of the disease. As specificity and prevalence decrease, so will the PPV, and many people with a positive test will in fact have false positive results. NPV are not substantially influenced by the prevalence of disease, in the range of prevalence estimates that can be found in most screening settings. However, a very high sensitivity will almost guarantee that no disease is missed and will translate into a high NPV.

1.1.1.4 Availability of an acceptable and safe screening test:

By definition, a screening test will be offered to symptom-free individuals. Most of the time, it will be offered to the general population, with the intention to reach a large number of individuals. It is thus imperative that screening tests have as few undesirable effects as possible, and they should not be associated with any significant morbidity.

1.1.1.5 Availability of valid diagnostic tests and treatments:

Screening tests do not always carry a definite diagnosis, but more often will identify individuals who should undergo more invasive diagnostic procedures [Miller 1985]. Treating patients on the basis of a screening test results alone may lead to considerable

over-treatment. Screening should not be undertaken if treatment options for the disease of interest are not available to patients.

1.1.1.6 Treatment of screen-detected disease should lead to a better outcome:

The first goal of screening is to lower the mortality associated with a given disease. This is achieved either by reducing incidence and/or by lowering the fatality rate. The incidence can be lowered if risk factors or precursor lesions are identified through screening and modified/treated. In some circumstances we may be satisfied with screening if it substantially lowers morbidity associated with the disease. It is sometimes taken for granted that treatment at an early stage of disease will lead to a better outcome. However, there are circumstances where this isn't so. For example, in a disease with early multi-system involvement making localized treatment ineffective, screening is unlikely to reduce mortality. If early treatment does not result in a better outcome, there is no reason to screen for this specific disease.

1.1.2 Ethical challenges of research on screening

When planning any research, ethical considerations must be considered from the outset. Although it is beyond the scope of this thesis to review general ethical principles of conducting research with human subjects, two specificities of research on screening interventions deserve at least a brief mention.

International organizations have established general principles that should guide the conduct of research involving human subjects. [Lebacqz 1986]. The Belmont report

[National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research, 1979] summarizes the ethical principles that should guide research as respect, beneficence and justice. The Council for International Organizations of Medical Sciences [CIOMS 2002] also built its Ethical Guidelines for Biomedical Research Involving Human Subjects around those 3 principles and the World Medical Association [WMA 2004] has expanded on them to develop its detailed Ethical Principles for Medical Research Involving Human Subjects, also known as the declaration of Helsinki.

Two of those fundamental principles are especially challenged in conducting research on screening interventions: the "pursuit of justice" (participants, bearing the inherent risks associated with an experiment, can also hope to benefit from future applications of the finding) and "beneficence" (first, do no harm). Although few participants may enjoy direct benefits in the course of a therapeutic trial, if a drug is found to be efficacious, sick participants can hope to eventually benefit from this new therapy. In contrast, most participants in a screening trial are healthy and will not develop the outcome (with or without screening), and so most of them will never derive any benefit from the proposed intervention. And more importantly, at least some patients will have a disadvantage: some will have false-positive results leading to unnecessary investigation, while others will be diagnosed and treated for a disease that would not have progressed. Is it then ethically justifiable to do such research?

If very carefully planned to minimize risk to participants and if there is true potential for net benefit, it may be, in order to further improve the health of a community [Adami

1994]. When doing so, the well-being of trial participants should always take precedence over potential benefits for society, as spelled out in the CIOMS guidelines. It is imperative to carefully balance the expected benefit to the community with the potential risks to patients [Miller 1985, Bailar 1983]. The screening procedures and the ensuing diagnostic procedures should be exceedingly safe, and adverse events should be aggressively monitored. Procedures used to obtain informed consent should clearly explain this situation. Interestingly similar concerns have been voiced over informed consent procedures before proceeding to screening in clinical practice [Lee 1993, Austoker 1999].

Fair and equitable allocation of scarce resources has surfaced as an important ethical challenge [Calman 1994, McKneally 1997]. Even outside the context of research, screening sometimes adds such a burden that an overextended health care system cannot support it. Human and technical resources may have to be spent on more urgent needs than the investigation of seemingly healthy individuals. The challenge is even more acute in the context of research. Indeed, the proper conduct of any trial entails time and energy. However, resource requirements can be especially daunting when considering a screening trial. Many therapeutic trials may answer their research question after enrolling hundreds of participants. When evaluating cancer screening in the general population, thousands (if not tens of thousands) of participants will need to be accrued. Given the limited human and material resources in the health care system, such trials may end up competing for resources needed for the treatment of already sick patients. Careful planning needs to ensure that such competition is kept to a minimum.

1.1.3 Evaluation of screening tests: design issues

Rigorous evaluation of screening tests and related activities ensure that the best options are identified prior to large scale implementation. Greenwald and al. have proposed a sequence of 5 research phases to "approach cancer control systematically", from hypothesis generation to methods development, followed by small controlled trials and larger population studies, and finally implantation and demonstration projects [Greenwald 1985]. In the context of this thesis, challenges associated with "defined population studies" are of particular interest.

A randomized controlled trial (RCT) will provide the best quality of evidence on the impact of a new test as a potential screening tool [Prorok 1984, Hennekens 1987, Woolf 1990, Fletcher 1996, Miller 2006]. Preliminary studies will have confirmed the inherent properties of the test (reliability, sensitivity, specificity) through smaller cross-sectional projects. The "ideal" RCT would then randomize participants from the target population to "usual care" vs. the screening intervention under study, as those interventions would truly be applied in clinical practice. Participants would then be followed long enough to verify if the new screening procedure is associated with a reduction in cause-specific (and overall) mortality [Mahon 2000, Miller 2006]. However, there are a number of difficulties in designing and implementing such a study: (1) some proof of principle must exist that the proposed intervention has a reasonable chance of reducing mortality, while at the same time, (2) there must remain a reasonable level of uncertainty as to the efficacy of the experimental screening intervention to ethically justify the randomization to a group with and a group without the intervention (state of clinical equipoise) [Shapiro 2000,

Freedman 1987]; (3) since the outcome is usually rare (such as cervical cancer), a very large number of participants need to be accrued in order to demonstrate an impact on cause-specific mortality; (4) if the study aims to identify a reduction in mortality, it may need a very long follow-up period, since there may be a long interval between screen detected disease and mortality from the disease. Given these conditions, it may be difficult to have a study that is unbiased and generalizable, yet feasible.

Different strategies can be used to circumvent the above-mentioned difficulties. If the absence of clinical equipoise prevents randomization (for example, if a screening test has been rapidly incorporated into practice and a consensus on its usefulness was reached within the clinical and scientific community before formal evaluation was carried out) a cohort design also referred to as "quasi-experimental studies" or demonstration projects may be used [Miller 1985]. This design follows screened and unscreened populations in order to assess differences in cause-specific mortality rates between the two populations. This can be achieved either by comparing geographical areas where screening is introduced to areas where it is not, or by comparing the same population before and after the introduction of screening, making sure all else is equal.

It can take years, even decades before a reduction in cause-specific mortality is apparent after the introduction of a screening test. To reduce the length of follow-up in an RCT, earlier outcome measures may be used to reach the conclusion that a screening test is useful, such as enhanced detection of precursor lesions, shift in stage to early cancers, increase in survival, or reduction in incidence of invasive cancer (which can be achieved

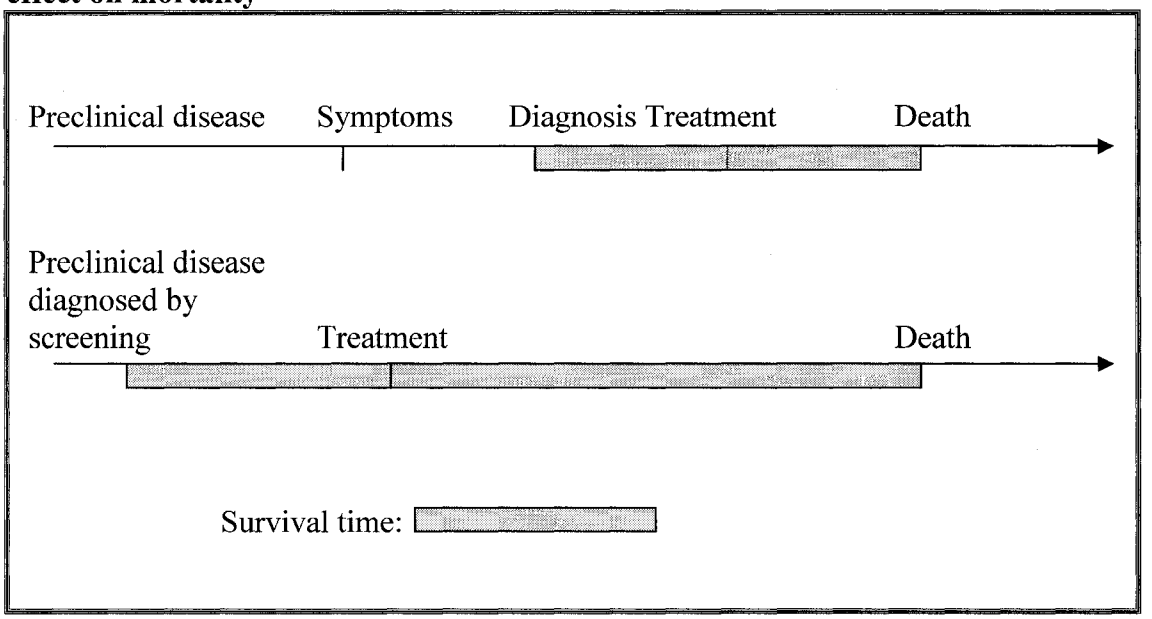
if the screening targets a precursor lesion). However, biases are more frequent with such measures and findings should be interpreted with caution [Miller 1985].

If the study population arises from the entire population where screening is offered, case-control studies can be viewed as efficient cohort studies, and present an alternative means to reduce follow-up. However, case-control studies face their own challenges, such as proper selection of cases and controls and correct definition of exposure [Cronin 1998, Miller 1985].

In any study evaluating the efficacy of screening, 2 types of biases are of particular concern: lead-time bias and length-time bias [Fletcher 1996]. Lead-time bias may be a problem in studies where the efficacy of a screening test is measured by comparing the survival experience of a group who is screened and another who isn't (Figure 2). By definition, survival time subsequent to a given diagnosis is the period between diagnosis and death. Survival time can thus be increased if death is postponed (which is what we are aiming for through screening), but also if diagnosis is made earlier in the course of the disease process. Indeed, in a case where early treatment is ineffective and does not alter the clinical course of the disease in question to the point of reducing risk of death, the group that has been screened may seem to have an increased survival, due only to earlier diagnosis.

Length time bias occurs when we screen for diseases that have heterogeneity in their course, with some cases being indolent, enabling the patient to live a long time with the

Figure 2. Lead time bias: how screening can increase survival time while having no effect on mortality



disease, and other cases being more clinically unfavourable, leading to death in a shorter period. At any point in time, there will be more patients alive who have the more indolent forms of the disease. When screening is done, these indolent cases are thus more likely to be detected. In a study comparing a group undergoing screening with another group where no screening is done, even if there is no benefit associated with screening, the fact that, on average, screening activities identify more indolent cases, the "screened" group will appear to have a more favourable outcome.

Given these challenges, it would be helpful to have consensus guidelines, such as those for the conduct and reporting of randomized controlled trials (CONSORT) and diagnostic studies (STARD), which have been published and updated [Bossuyt 2003, Moher 2001]. Even if some guidance can be found in those publications, CONSORT addresses mainly issues relating to the evaluation of therapeutic measures, which are not all relevant to

screening, at the same time omitting key features that need to be addressed when evaluating screening. On the other hand the STARD statement focuses only on study features that will provide unbiased evaluation of test validity, without addressing the other aspects that make a good screening test (correct choice of outcome, length of follow-up period).

Unfortunately, given the challenges outlined above, many studies relating to screening tests are not of optimal methodological quality. Studies on new screening tests for cervical cancer being no exception, most systematic reviews have not been able to provide firm conclusions on the value of these new tests, mainly because of the lack of good-quality studies [McCrory1999, Broadstock 2000, Karnon 2004, Noorani 2003].

1.2 Screening for cervical cancer

After discussing principles relating to screening in general, I will focus on the specific topic of cervical cancer screening. I will review why it is justified to screen for cervical cancer, what is the current strategy, and the strengths and limitations of this strategy.

1.2.1 Justification for cervical cancer screening

1.2.1.1 Burden of disease

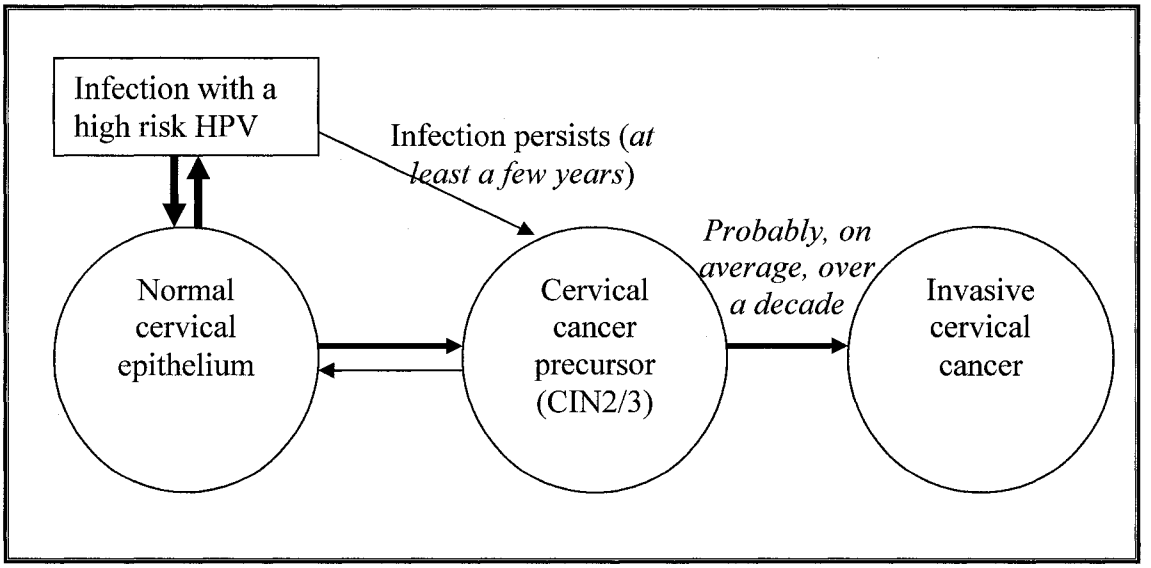
Worldwide, cervical cancer remains the second most frequently diagnosed cancer in women and incidence rates peak as high as 87/100,000 women in certain African and Caribbean areas [Parkin 2002]. In Canada, it is estimated that 1350 new cervical cancer

cases will be diagnosed in 2007, and that this disease will claim 390 lives [CCS 2007]. Although this fact places cervical cancer in 13th place in terms of cancer incidence and 15th in terms of cancer mortality in Canadian women, many consider these figures still too high, considering that the disease is almost entirely preventable. Moreover, contrary to more frequent cancers which strike mainly older women, cervical cancer strikes mostly women in their forties and fifties, which has an important effect in terms of lost years of life and a devastating effect on the social network of communities [Ries 2007, Davies 2005].

1.2.1.2 Natural history of the disease

The pre-invasive stage of squamous cervical cancer is probably the best described of all cancers [Richart 1968, Peterson 1956, Ostor 1993, Pinto 2000, Schiffman 2003, Schiffman 2007]. Cervical cancer and its precursors are caused by the infection of the cervical epithelium by oncogenic types of human papillomaviruses (HPVs) [Zur Hausen 1991, Munoz 1992, Schiffman 1993, Bosch 1995, IARC 1995, Franco 1999a, Walboomers 1999, Bosch 2002, Trottier 2006]. Over a 100 different types of HPVs have been identified (and given a number) and approximately 40 infect the genital area. Of those, 15 to 20 types have been linked to cervical cancer, and are referred to as "high risk" or "oncogenic" types [Munoz 2003, Trottier 2006, Schiffman 2007]. The products of two oncogenes of those high risk types, E6 and E7, interfere with the human cells' mechanisms which regulate cell proliferation and DNA repair, leading to unregulated cell growth, genomic instability, and under certain conditions, malignancy [Baldwin 2003, Franco 2005].

Figure 3. Natural history of HPV infections, cervical cancer precursors, and cancer

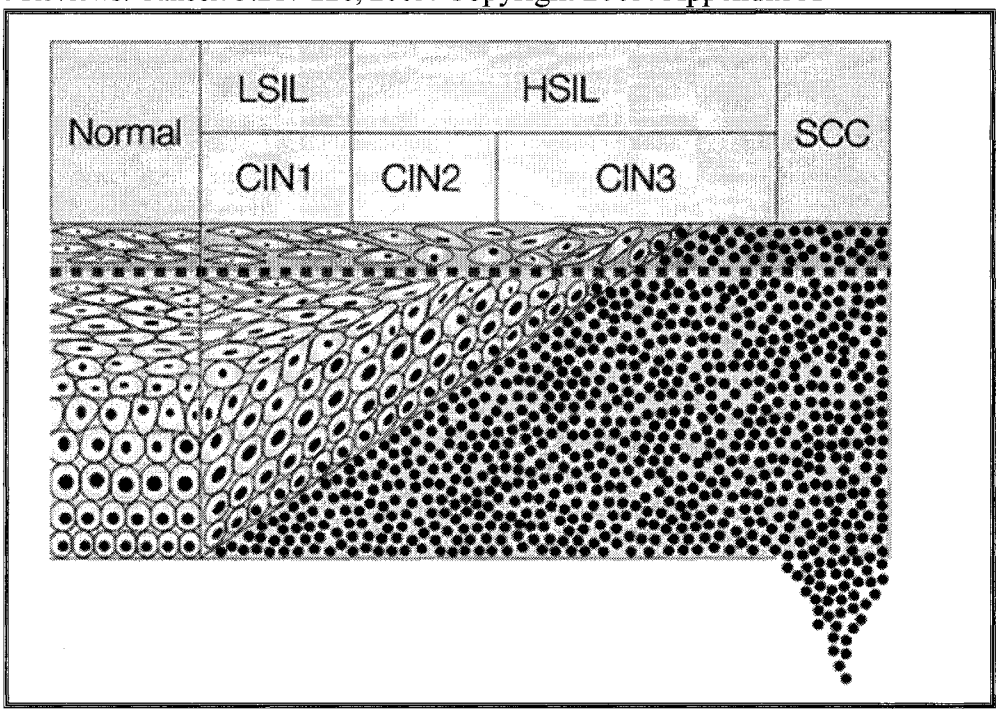


It is estimated that 80% of sexually active women will be infected by a genital HPV at some point in their lifetime, most of them in the first years following initiation of sexual activity [Syrjänen 1990, Bosch 2006]. Most will clear the infection (even by an oncogenic type) without noticeable sequels [Hildesheim 1994, Ho 1998b, Moscicki 1998, Franco 1999a, Molano 2003]. If the infection by an oncogenic type is not cleared and persists for a prolonged period in the cervical epithelium, preneoplasic changes can occur [Koutsky 1992, Ho 1995, Remmink 1995, Chua 1996, Wallin 1999, Schlecht 2001, Kjaer 2002] (Figure 3).

Precursor lesions (cervical intra-epithelial lesions or CIN) are divided in 3 categories (CIN1, CIN2 and CIN3) based on their histological aspect (Figure 4). CIN1 is the less severe category, with only a third of the thickness of the epithelium showing precancerous changes. CIN1, representing simple HPV infection, often regresses without treatment. CIN3, where more than two thirds and up to the full thickness of the epithelium shows precancerous alterations, carries the highest risk of progression to the invasive stage [Moscicki 2004, Saslow 2007]. The pre-clinical phase of CIN3 probably

lasts from a few years to as much as 20 years before invasion occurs [Hidelsheim 1999, Saslow 2007]. This leaves enough time to find the precursor lesions with a screening test. Although CIN3 may be the ideal stage for intervention, CIN2, a condition intermediate to CIN1 and 3, is often selected as the diagnosis target in clinical practice to offer a margin of safety in the identification of CIN3 [Wright 2003a, Castle 2007]. CIN2 and CIN3 are often referred to as" high grade" cervical cancer precursors or lesions. Contrary to late stage cervical carcinomas, intra-epithelial lesions are often asymptomatic and can only be identified by screening. Because the identification and treatment of high grade lesions halt the natural progression of the disease towards invasion, screening for cervical cancer precursors has had an impact on both incidence and mortality from the disease [Franco 2002].

Figure 4. Schematic representation of cervical intra-epithelial neoplasia
(Reprinted with permission from Macmillan Publishers Ltd: Nature reviews.Cancer. Baldwin P. Laskey R. Coleman N. Translational approaches to improving cervical screening. Nature Reviews. Cancer. 3:217-226, 2003. Copyright 2003. Appendix A



1.2.1.3 Availability of a valid screening test

The cervical cytology, or Pap test as it is most commonly known, has been the central test used for screening for cervical cancer in the last decades. This screening test was discovered "by accident" by a Greek pathologist, Dr Papanicolaou (hence the name Pap test). Dr Papanicolaou's initial interest in studying vaginal cytology was to describe the cellular changes relating to the menstrual cycle. However, he eventually identified cellular changes characteristic of cancer. His first landmark paper in 1928 was essentially ignored. He later joined Dr Traut, a gynecologist who was convinced of the importance of his findings. Together they were able to demonstrate how cytology could identify cervical cancer in its preclinical phase [Papanicolaou 1943]. This publication was a turning point in the management of cervical cancer, and in the following decade the Pap test became widely accepted as a routine screening test.

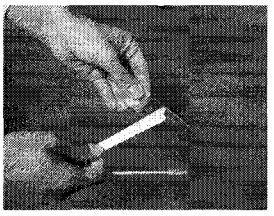
The conventional Pap test is performed as follows. The cytology samples are collected by scraping a wooden spatula (or another similar device) against the cervix, under direct visualisation of the transformation zone, the specific area of the cervix where preneoplasic lesions arise. The collected material is smeared onto a glass slide and sprayed with a fixative to preserve cell morphology. At the cytology laboratory, samples are then stained and evaluated by a cytotechnician for signs of cellular changes consistent with CIN (Figure 5). Much effort has been devoted to the standardisation of terminology for reporting results of cytology. In North American and most parts of Europe, the Bethesda 2001 classification is used to classify smears (Table 2).

Figure 5. Steps of conventional Pap test collection, preparation and interpretation
(Pictures used with permission from N Chamberlain and Louisiana State University Health Sciences Center. Appendix A)

Step 1.

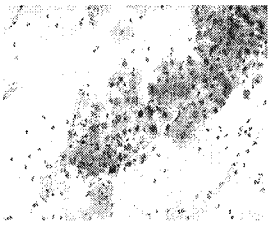
The cervical sample is collected by scraping a wooden spatula against the cervix and rotating a brush in the cervical canal.

Step 2.



The collected material is smeared onto a glass slide and sprayed with a fixative

Step 3.



The material on the slide is stained and evaluated by a cytotechnician for signs of cellular changes consistent with CIN

Pap cytology is credited with having contributed to a reduction of as much as 75% in the mortality and morbidity associated with cervical cancer during the past 40 years [Morrison 1992, Liu 2001, IARC 2005, Anttilla 2004a, Franco 2005]. However, there have been no prospective controlled trials of Pap screening efficacy, either randomized or not. The evidence for the efficacy of Pap smear screening in cervical cancer comes from case-control and ecologic studies, which indicate that the risk of cervical cancer is greater

Table 2. Cytological classification abbreviations, according to the Bethesda 2001 terminology [Solomon 2002].

Cellular	Abbreviation	Signification
Squamous	ASC	Abnormal squamous cells
	ASC-US	Abnormal squamous cells of undetermined significance
	ASC-H	Abnormal squamous cells, cannot exclude a high grade lesion
	LSIL	Low grade squamous intra-epithelial lesion (Human Papillomavirus infection, CIN1)
	HSIL	High grade squamous intra-epithelial lesion (moderate and severe dysplasia, carcinoma in situ; CIN 2 and CIN 3)
	Squamous cell carcinoma	Squamous cell carcinoma
Glandular	AGC	Abnormal glandular cells
	AIS	Adenocarcinoma in situ
	Adenocarcinoma	Adenocarcinoma

in women who have not been screened, with time since last normal smear, or with lower frequency of screening; and that cervical cancer incidence and mortality rates have decreased following the introduction of cytology screening in Scandinavian countries, in the United Kingdom, in Canada, and in the US when adequate coverage was achieved [Petterson 1985, Quinn 1999, Franco 2002].

Despite its success in reducing incidence and mortality associated with cervical cancer, the Pap cytology has important limitations. Nanda et al, have published an extensive review on the subject [Nanda 2000]. After evaluating over a thousand studies, the authors concluded that only 94 studies focusing on the performance of the conventional Pap smear could be included in the review. Only 12 of those studies focused on patients

undergoing screening and verified all screening test results or a random sample of participants with histology and/or colposcopy. In the 6 studies for which data was available on the performance of the Pap test to detect CIN2 or worse, some sensitivity estimates were as low as 44% and a few as high as 90%, and the specificity ranged from 91% to 98%. This review illustrated that even if Pap testing was widely used there was a paucity of data regarding its accuracy in cervical cancer screening in the general population, since only 6 studies focused on the relevant population and had some verification of disease status to allowed the calculation of sensitivity, specificities and predictive values. It also highlighted the fact that the sensitivity of the conventional Pap smear was lower than what was previously believed and that its performance varied substantially across settings, putting its reproducibility into question. Given those characteristics, the Pap test needs to be repeated frequently in order to provide adequate protection against cancer, adding to the cost and overall burden to the health care system. It is possible that the Pap smear would not have been introduced into practice had it been submitted to the phases of research proposed by Greenwald [1985], since it would probably have shown disappointing results in methods development and/or small controlled trials. However, despite its limitation, the Pap test remains a cancer control success story because, ultimately, its impact on cervical cancer mortality is undisputed.

However, the realization of its shortcomings (the wide range of performance and its low sensitivity in many laboratories) prompted the evaluation of alternative or complementary tests. A more sensitive test could potentially not only prevent more cases of invasive cancer, but also improve the efficiency of the screening process. Indeed, cervical cancer is probably the only cancer for which some professional organisations have recommended

up to 50 rounds of screening in a lifetime [Parboosingh 1999, Miller 2002, Smith 2006]. Given the natural history of the disease, screening intervals could be safely lengthened if a more sensitive screening test were available.

1.2.1.4 Safe and acceptable screening test

There is no doubt that the Pap test is safe. It has no inherent risk or complications. It has now been part of medical routine for so long that many women in western countries take for granted that a pelvic exam, with a Pap test, is part of routine medical care [Anhang 2005]. However, this is not true across all countries, ethno-cultural groups or socio-economic classes, which has lead to the evaluation of alternative screening modalities where women can obtain the sample themselves [Dzuba 2002, Nobbenhuis 2002, Belinson 2003, Dannecker 2004, Forrest 2004, Anhang 2005, Baldwin 2005, Ogilvie 2005, Bais 2007]. Because the Pap test needs to sample a very specific region of the cervix, self sampling is not an option.

1.2.1.5 Availability of diagnostic tests and treatments

The colposcopic examination forms the basis of the diagnostic procedure for cervical cancer precursor lesions. It entails that an experienced examiner looks at the cervix through a magnifying lens with proper illumination, after the application of acetic acid and/or iodine solution(s). Specific patterns of epithelium change can be identified as possible CIN. Histological examination of colposcopy-directed biopsies of such lesions is the accepted method to obtain definite diagnosis following an abnormal Pap smear [Wright 2002]. It is widely available in resource-rich countries. Treatment protocols,

aimed at removing or destroying the transformation zone are well described [Wright 2003a]. Cryotherapy, laser fulguration or ablation, cold-knife conisation and loop electro-excision procedure (LEEP) all have success rates around 90%, and complication rates below 10% [Kolstad 1976, Loizzi 1992, Mitchell 1998b, Martin-Hirsch 2005]. Their impact on reproductive health is minimal if the recommended depth of treatment is not exceeded [Sadler 2004b, Acharya 2005, Kyrgiou 2006].

1.2.1.6 Treatment of screen-detected disease should lead to a better outcome

If screening is done regularly, cervical lesions are usually identified at the pre-cancerous stage, when the success of treatment is the norm and the morbidity associated with outpatient treatment is limited. Early stage cervical cancer (stage 1 and early stage 2) can also be asymptomatic and identified by screening. Although treatments are then more invasive (radical surgery and/or radiotherapy), the prognostic remains very good: 90% disease free survival at 5 years. This is in stark contrast to lesions identified after often severe symptoms have lead to a diagnostic work-up, leading to a diagnosis of stage 3 or 4 invasive cancer. Under such circumstances, treatments are associated with extreme morbidity and the 5-year survival rate hovers around 30% [Ries 2007].

1.2.2 Current prevention strategies for cervical cancer: the strengths and limitations

1.2.2.1 Successful chain of events following screening

In order to fully appreciate the strengths and limitations of the current screening process, it is imperative to look at screening as a chain of events, where all the links in the chain have to be successfully completed in order to have the desired outcome: prevention of

cervical cancer, and prevention of death from cervical cancer. The steps that need to be completed in succession are the following: (1) women need to receive the screening test, (2) the screening test needs to correctly identify significant lesions, (3) women with abnormal screening tests results need to be investigated with the correct diagnostic procedures, (4) women confirmed to have lesions should be treated appropriately.

It has been shown that among the cervical cancer cases diagnosed in resource-rich settings where there are no (or incomplete) organized screening programs, approximately half are due to a lack of screening, and the other half to errors in the screening process: false negative screening test, lack of referral to colposcopy after an abnormal screening test, errors in diagnostic or treatment of precancerous lesions [Stuart 1997, Sadler 2004a, Leyden 2005, Spence 2007]. Given these pitfalls, it is evident that cervical cancer screening research and implementation efforts should not only focus on devising a better screening test but also on screening delivery.

1.2.2.2 Screening delivery

Screening for cervical cancer can essentially be delivered in 2 different ways: in the context of an organized program, or by an opportunistic approach. The opportunistic approach relies on women to go to their health care professional for screening, or for health care professionals to offer screening to women who visit their office for other health matters. Organized screening activities take place within a service delivery structure that: (1) identifies women who should be screened (most frequently based on age groups) from central registries and invites them for screening, (2) respects recommended screening intervals, (3) arranges follow-up for abnormal screening and

diagnostic results, (4) maintains quality assurance programs, (5) evaluates the different steps of the screening process and implements correctives when needed, (6) is active in health promotion activities [Miles 2004, Howlett 2006]. The organized approach has been fully implemented in Nordic countries, and at least some aspects have been implemented in most European countries [Anttila 2004]. Opportunistic screening often results in over-screening of certain sub-groups of the population, not surprisingly those at low risk, and under-screening of other groups, sometimes those at highest risk [Bastani 2002, Cyrus-David 2002, Selvin 2003, Abraido-Lanza 2004, Anttilla 2004, Miles 2004, Breen 2005, Saint 2005]. Overall, it may be possible to achieve similar gains in terms of controlling mortality at the population level through opportunistic and organized delivery, but organized programs increase the efficiency of the screening process [Nieminen 1999, Hanselaar 2002, Miller 2002, Anttilla 2004, Miles 2004, Ronco 2005]. For example, countries with an organized screening program will usually be able to screen at less frequent intervals and stop screening at an earlier age while achieving the same level of control on incidence and mortality [Van Ballegooijen 2000].

1.2.2.3 The current state of cervical cancer screening in Canada

Although organized screening is desirable, it has not been fully implemented in Canada, probably because of the initial increase in costs and workload associated with such a program. Some provinces have started implementing at least certain aspects of an organized program. Some have almost completed the process (British Columbia) while others have yet to start (Quebec and New Brunswick).

The Pap test is used in all provinces for cervical cancer screening. Most public laboratories still use the conventional Pap test. In Ontario, a commercial semi-automated liquid-based cytology system has been used in the past 4 years. Recommended screening intervals vary from yearly to every 3 years. For provinces with available data, we know that between 63% and 75% of women between the ages of 20 and 69 have had a Pap smear in the 3-year period ending in 2003 [Rose 2006]. Colposcopy services are available and accessible. Canadian colposcopists evaluate and treat cervical lesions in accordance with the guidelines of the American Society for Colposcopy and Cervical Pathology (ASCCP) [Wright 2002, Wright 2003a], as detailed in sections 1.2.1.5 and 1.2.1.6.

Overall, Canada is recognized has having achieved a very good control of cervical cancer mortality, with an average incidence rate of 8/100,000 women. However, this obscures the fact that success is uneven and certain regions have age-standardized incidence rates that are almost double (11/100,000 vs. 6/100,000), and mortality rates that are four times (4/100,000 vs. 1/100,000), that of more successful regions [CCS 2007]. Moreover, to achieve such success in the context of opportunistic screening using a suboptimal screening test, over-screening of certain segments of the population compensates for the under-screening of others. A just allocation of resources and efficiency of the process remain key challenges. The following section will review options which may improve the current strategy for cervical cancer prevention.

1.2.3 Moving forward: what are our options

1.2.3.1 Primary prevention

Prevention and control of genital HPV infection through sexual behaviour modification could, in theory, contribute to the prevention of cervical cancer. Unfortunately, interventions aimed at a modification of sexual practices exert only a modest effect of short duration, and there is limited data showing that these interventions ultimately result in a reduction of sexually transmitted infection (STI) incidence [Sheperd 2005]. Moreover, most interventions encouraging safer sexual lifestyles focus on promoting condom use, whose role in preventing HPV transmission remains controversial [Manhart 2002, Vaccarella 2006]. A recent study designed specifically to address this issue was able to show a 70% reduction in risk of acquiring an HPV infection when condoms were used consistently. However, even in this group of college educated women using condoms for all instances of intercourse over an eight month period, the incidence of HPV infection was still 38 per woman-year [Winer 2006]. Given that most sexually active men and women will be infected by HPV at some point in their lifetime (section 1.2.1.2) and that sustained modification of sexual behaviour is difficult to achieve, it is doubtful that such interventions will play a significant part in the prevention of cervical cancer.

Vaccination, on the other hand, offers the best hope of eventually controlling cervical cancer through primary prevention. Even in settings with effective screening strategies in place, vaccination combined with screening will probably further improve cervical cancer control [Kulasingam 2003, Goldie 2004]. One vaccine (against HPVs 6, 11, 16, 18) is

already approved for use in Canada, while another (against HPVs 16 and 18) is expected to be available in the short term. Both vaccines use "virus like particles" (VLP) and are prophylactic. The results of Phase II and Phase III clinical trials are extremely promising. They indicate that these HPV vaccines are over 90% effective in preventing: (1) acquisition of new infections, (2) development of persistent HPV infections, and (3) high grade cancer precursors related to cancer types included in the vaccine in women not previously exposed to those types [Harper 2004, Villa 2005, Harper 2006, FUTURE 2007, Garland 2007, Villa 2006, Paavonen 2007].

There is still much to be learned however, before these gains can translated into actual reduction in cervical cancer incidence. Not all women will be vaccinated and those already exposed to HPV 16 or 18 may not benefit as much from vaccination [Koutsky 2006, FUTURE 2007, Sawaya 2007, Saslow 2007]. Evidence from controlled trials has been obtained for monovalent (HPV 16, oncogenic), bivalent (HPVs 16 and 18, both oncogenic), and quadrivalent (HPV 6 and 11, non oncogenic; and HPV 16 and 18) vaccines. A future vaccine will have to include more high-risk HPV types to be largely effective against cervical cancer [Sawaya 2007]. A vaccine targeting the 7 most common oncogenic types could prevent as much as 87% of cervical cancers worldwide [Munoz 2004]. The fact that, for now, only two of the cancer-causing HPV types are included in the vaccines makes it mandatory to continue screening.

1.2.3.2 Screening

Before the full benefits of vaccination are widely available, screening remains our best option to prevent cervical cancer. Therefore, we need to address the principal

shortcomings of our screening strategy: inadequate coverage and inadequate validity of the screening test.

Research and evaluation of screening delivery have shown that organized screening is both more efficacious and more cost-effective than opportunistic screening (section 1.2.2.2). They contribute to increase coverage and decrease errors in follow-up. Unfortunately, given the current state of public health care systems, it has been difficult to fund all aspects of organized programs.

Ultimately, it is possible that the work to identify a better screening test will eventually help increase coverage. Indeed, if a more sensitive test is available, women and their health care providers may feel more secure in lengthening the screening interval, thus reducing the frequency and ultimately the number of screening tests. Cervical cancer screening activities based on a screening test other than the Pap test may even prove to be less expensive on the long run. Resources could be better spent reaching under screened populations and putting in place an organized program. Below is a review of options for better screening tests.

Make the conventional Pap test better. To ensure the validity of the test, health care providers should be properly trained in taking Pap smears. The cervix should be entirely visualized. Excess mucus should be wiped off and the transformation zone (the specific area of the cervix where precancers arise) should be sampled. After proper preparation, the slide should be examined by an experienced cytotechnician in a laboratory applying adequate quality control measures. Identification of preneoplasic changes is highly

subjective, and the repetitive nature of the work increases interpretation errors [Cuzick 2006a]. Because barriers to improvement are present at each of those steps, even organized programs with quality assurance measures have found it very difficult to improve the performance of the conventional Pap test [Shaw 2002].

Liquid based cytology. Liquid based cytology (LBC) represents an alternative to conventional Pap, but only in terms of slide preparation. The cervical sample is obtained in much the same way, but instead of being transferred directly on a glass slide, it is suspended in a liquid medium. At the laboratory, the sample is vortexed, which allows breakdown of mucus and uniform cell dispersion. An automated device then transfers a sample of cells on a glass slide as a thin layer. This process makes it possible to have less blood and inflammatory cells and the epithelial cells more evenly spaced on the slide. The subsequent staining and reading is similar to what is done with the conventional Pap smear. There was initially considerable enthusiasm, as most studies pointed to an improvement in sensitivity [Austin 1998], with the result that LBC was quickly incorporated in practice in the US, where litigation is of particular concern [Saint 2005].

However, when focusing on studies designed to compare the accuracy of the liquid-based technology to that of the conventional Pap, it became evident that most of them have severe methodological limitations, so as to make their conclusions less than definitive. Some studies were done on a high risk-population, selected to undergo colposcopy, making generalization of their findings to an average risk population difficult. Given the novelty of LBC, cytotechnicians had to be trained in interpreting LBC; this recent training alone could have contributed to increased accuracy in diagnosis. Most studies used a split

sample design; meaning that only one cervical sample was collected. A conventional smear was prepared, and the residual material was then put in a liquid medium. It was thought at first that this would disfavour LBC, because fewer cells would be available for analysis; however, it is possible that most unwanted material got transferred on the first slide, with the conventional Pap having more mucus, blood and inflammatory cells [Obwegeser 2001]. Indeed, in their meta-analysis, Noorami et al, found that only studies using the split sample design showed a sensitivity advantage for LBC [Noorani 2003]. Most studies did not include any verification of diagnosis by colposcopy and/or histology. It is then only possible to compute relative (rather than absolute) screening indices. For example, McCrory et al, in their review, identified only one study on LBC that had sufficient data on histological diagnosis to compute unbiased sensitivity and specificity estimates, which were similar to estimates for conventional Pap [McCrory 1999]. Given these shortcomings, other countries have moved more slowly before incorporating LBC in standard screening practices.

In recent years, national health technology assessment bodies in Canada, USA, Australia, New Zealand and Europe have reviewed thoroughly the available evidence for the efficacy and cost effectiveness of LBC [McCrory 1999, Broadstock, 2000, Hanselaar 2002, Moss 2003, NICE 2003, Noorani 2003, Karnon, 2004, Davey 2006, Hulstaert 2006]. They have each confirmed that the quality of the available evidence to compare the value of LBC to that of the conventional Pap is poor. As for sensitivity and specificity estimates, they have concluded either that the quality of the evidence precludes any conclusion or that LBC provides, at best, a very modest advantage in terms of sensitivity when biopsy-proven high-grade lesion is the selected outcome. A recent RCT including

over 40 000 participants could not identify an advantage in terms of sensitivity when comparing LBC to Pap [Ronco 2007b]. With sensitivity similar to that of the conventional Pap, it is unlikely that the introduction of LBC would significantly impact on cervical cancer control.

HPV testing. Ever since oncogenic HPVs have been demonstrated to be the central causal factor for cervical cancer, there has been considerable interest in testing for this virus as a screening modality. A better understanding of the natural history of HPV infection and CIN and the development of new technologies now make it possible to incorporate HPV testing in screening programs. Polymerase Chain Reaction (PCR) protocols, which are based on the amplification of DNA, have been considered the gold standard for etiologic research purposes. PCR has the ability to distinguish among individual HPV types and has a low threshold for HPV detection. However it is exactly this low threshold that makes it less suitable for screening purposes, identifying minute amounts of HPV with no clinical correlate.

New technologies were developed to design a test more suitable than PCR for mass screening. The first generation of assays specifically designed for large scale use lacked the necessary accuracy to be useful [Franco 2003], but improvements in technology have made it possible to consider their use for screening purposes [Hall 1996]. There is currently only one commercially available test that is well suited for this purpose since it has been calibrated with the specific goal of detecting cervical lesions: The Hybrid Capture2 test (HC2) (Digene Inc., Gaithersburg, Maryland, USA) . This test uses nucleic acid hybridization and subsequent signal amplification (Figure 6). It can detect one or

more of 13 high risk HPV types: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68. The manufacturer recommends using 1pg DNA/ml (equivalent to 5000 viral copies) as the positivity threshold. After initial proof of principle studies pointed to a possible advantage in terms of sensitivity of HPV testing for the detection of high grade lesion, controlled studies have been carried out [Cuzick 1999a, Kuhn 2000, Ratnam 2000, Schiffman 2000, Belinson 2001, Blumenthal 2001, Clavel 2001]. From these trials, there is strong evidence that HPV testing is more sensitive than Pap cytology to detect CIN2 or worse lesions. However, certain limitations of the study designs precluded the adoption of HPV testing in cervical cancer screening (reviewed in section 1.3).

Figure 6. Hybrid Capture 2 Technology *(© Digene Corporation – used by permission. Appendix A)*

Step 1. Formation of Hybrids

The HC2 vial contains RNA probes targeting oncogenic HPV. If such DNA is present in the patient sample, it will bind with the RNA probes and form hybrids

Step 2. Capture of Hybrids

Antibodies targeting RNA-DNA hybrids coat the vial. The antibodies will "capture the hybrids". These will stay in the vial when the specimen is washed.

Step 3. Signal Amplification

Other antibodies are then added to the vial. These antibodies not only bind to hybrids but also emit light through a chemiluminescent reaction when they bind. The emission of light is measured by a luminometer and is proportional to the amount of oncogenic HPV present in the sample.

It should be noted that HPV testing has been shown to be both efficacious and cost-effective in the triage of equivocal cytological abnormalities [Cox 1995, Manos 1999, Solomon 2001, Zielinski 2001, Kim 2002, ALTS 2003, Arbyn 2004, Bais 2005, Berkhoff 2006, Kulasingam 2006, Ronco 2007a]. In a triage context, women have a Pap test (conventional or liquid-based) first and only women with equivocal results (ASC-US) will have an HPV test. Women with a negative HPV test result only need to be re-screened by cytology after 12 months, while women with a positive result are referred for diagnostic evaluation [Wright 2002]. To clearly differentiate when referring to HPV testing as a triage tool as opposed to HPV testing used as a first-line screening test, the latter is often referred to "HPV testing in primary screening".

1.3 Comparing Pap and HPV testing for primary screening of cervical cancer: review of the literature

This research project was initiated in 2001. The review of studies comparing Pap to HPV testing for primary screening of cervical cancer published up to 2001 (Table 3) provides an understanding of design and analysis issues that needed to be addressed when planning this project in order to overcome limitations of available data (a complete update of the available evidence is presented in Table 26). These issues are detailed in the following sections. It should be noted that Table 3 includes all studies which reported results of screening of women **aged 30 or older**. Indeed, the results of the first studies comparing Pap to HPV testing typically included women 18 years or older. However, it was shown that HPV testing was not adequate for younger women because viral infections in this age group are more likely to be of a transient nature than those in older women, and not

associated with significant disease [Hildesheim 1994, Moscicki 1998]. Restricting HPV testing to older women improved the specificity of HPV testing [Cuzick 1995, Cuzick 1999a, Schiffman 2000, Ratnam 2000]. HPV testing is thus not recommended for women younger than 30 [Wright 2004].

Table 3. Studies comparing Pap cytology and HPV testing with HC2 in primary screening for cervical cancer and its precursor lesions up to 2001[1]

Study	Country	Size	Age	Sensitivity		Specificity		Referrals	
				HPV	Pap	HPV	Pap	HPV	Pap
Belinson, 2001	China	1997	30-45	95	94	85	78	18	25
Cuzick, 1999a	U.K.	2988	≥35	95	86	na*	90	7	6
Kuhn, 2000	South Africa	2944	35-65	88	78	82	97	22	8
Ratnam, 2000 [2]	Canada	595	>30-35	82	40	94	97	5	4
Schiffman, 2000[3]	Costa Rica	3536	>40	93	na	94	na	7	na
		8554	>18	88	78	89	94	12	7
Blumenthal, 2001	Zimbabwe	2073	na	80	44	61	91	43	13
Clavel, 2001[4]	France	5671	>30	100	75	89	95	12	6

[1] When more than one article reported results on the same study population, the most complete report was used. When reports included different results, the following were selected: population of women older than 30; HC2 HPV test format at a positivity threshold of 1pg/ml; Pap positivity threshold of ASC. [2] screening indices for women over 30; referrals for women over 35.
[3] Age stratified results only available for HPV testing ;
[4] Conventional Pap results and LBC results combined

1.3.1 Design

All studies were **cross-sectional**: participants received both a Pap and an HPV test at enrolment and were examined by the gold standard (colposcopy) either immediately or shortly after the screening test results were available. The **Pap** test was always collected **first**. The remaining material on the sampling device was then used for HPV analysis, or

a second sample was collected form the cervix and analyzed to detect the presence of HPV. This may have contributed to "sampling interference", since all biological material analyzed for HPV content was collected as a second sample. There is evidence that for certain types of screening tests, sampling order may affect the performance of the test [Obwegeser 2001].

1.3.2 Populations

The progression of CIN is known to be heterogeneous, with some lesions persisting as low grade or high grade precursors while others progress more rapidly to cancer [Hildesheim 1999, Saslow 2007]. For this reason, intensity of screening in a population will affect the underlying spectrum of "aggressiveness" of cervical lesions. Indeed, if a screening test is evaluated in a never screened population most lesions identified will be prevalent, more indolent ones. On the other hand, in settings where most women are screened annually and identified lesions treated appropriately, each round of screening will identify mostly incident lesions. The difference in disease spectrum is a recognized factor that may affect the sensitivity and specificity of a screening test. Since more than half of studies in Table 3 were conducted in **never screened populations** [Belinson 2001, Khun, 2000, Schiffman 2000, Blumenthal 2001], it appeared important to re-assess screening test performance in our own setting.

Compounding this problem of different disease spectrum is the fact that 2 of the studies in never screened populations [Khun 2000, Blumenthal 2001] may have included an important proportion of **HIV positive women**. The natural history of HPV infection and the course of cervical cancer are known to be affected by HIV induced

immunosuppression [Palefsky 2001]. This may have contributed further to differences in the underlying spectrum of cervical disease.

1.3.3 Screening tests

The conventional Pap cytology was used in most studies. However, this test was read in **expert laboratories** if adequate cytology facilities did not exist, which may have given an optimistic appraisal of Pap performance. The Hybrid Capture system was first available as a **tube test (HCT)** with a lower analytical sensitivity, which was not deemed appropriate for primary screening, and eventually was replaced by the HC2. Only 3 studies in Table 3 used the HC2 test for the analysis of all study samples [Belinson 2001, Blumenthal 2001, Clavel 2001]. Ratnam and al. switched to HC2 when it became available, but did not account for this in the analysis. Cuzick, Kuhn and Schiffman (Table 3) retested a proportion of samples with the HC2 test and used those results to estimate screening indices for the whole study population.

1.3.4 Diagnostic procedures and outcome definition

Colposcopy was used for disease diagnosis, but protocols varied. Most studies did not systematically perform biopsies and only used **colposcopic impression as the gold standard** for diagnostic (all but Belinson 2001). This may have lead to serious misclassification of disease status as colposcopic impression alone has been shown to be unreliable to establish presence or absence of high grade intra-epithelial lesions [Hopman 1995, Mitchell 1998a, Massad 2002, Guido 2003, Sideri 2004, Elfgren 2005]. More importantly, most studies did not look for disease in women with negative screening test

results, assuming that they were truly free of disease [Cuzick 1999a, Khun 2000, Clavel 2001]. Those who assessed a random sample of women with negative screening tests, assessed such a small sample that the absence of disease in this group could not be ruled out, given the expected low prevalence: Ratnam found no CIN2 or worse lesion in 145 women, and Schiffman found none in 128 women.

1.3.5 Analysis

It is important to note that the sensitivity and specificity estimates of most studies shown in Table 3 are relative, not absolute, **because of verification bias.** Such a bias originates from the fact that often only participants with positive test results undergo disease verification. Since the case status of participants with negative screening test results is not verified, sensitivity estimates are typically overestimated. Unbiased estimation of screening indices must take into account the prevalence of disease among those who are test negative, which in theory would require disease verification among all subjects. For cervical cancer screening studies, obtaining a biopsy for histological verification in all participants without positive tests is difficult to realize because of ethical (the discomfort associated with the procedure) and practical considerations (limited availability of diagnostic resources and costs incurred). Studies which do not address this bias relied on the fact that with two or more tests there were always combinations of either Pap negative or HPV negative women with verified disease status available for analysis. However, the biasing effects of the **unequal verification of disease status depending on screening test results** can be strong and may lead to estimates of screening efficacy that cannot be generalized for cost considerations and other public health uses [Franco 2000; Ratnam 2000] (Table 4). Only the estimates of Belinson [2001] and Blumenthal [2001] (by

design, all women underwent colposcopy and biopsy) and Ratnam [2000] (by analysis, data on a random sample of screen-negative women were used to correct for verification bias) were not affected by verification bias (although Schiffman performed colposcopy on a limited number of test negative women, the results were not corrected for verification bias in the analysis).

Although it has been suggested that the precision of estimates of diagnostic accuracy should always be reported [Harper 1999], most studies in Table 3 reported only point estimates of performance indicators. Given the relatively small sample sizes, the confidence intervals of Pap and HPV testing performance indicators would have largely overlapped.

Table 4. Illustration of impact of verification bias on test performance estimates

Diagnostic ascertainment	% with diagnostic ascertainment	Screening test result	Disease present	Disease absent	Disease unknown
(1) Complete	100	Positive	25	48	0
	100	Negative	25	902	0
(2) Incomplete	80	Positive	20	38	15
	10	Negative	3	90	834

Screening Parameter	Correct value (%) (complete diagnostic ascertainment)	Biased value (%) (incomplete disease ascertainment)
Sensitivity	50	87
Specificity	95	70
Positive Predictive Value	34	34
Negative predictive Value	97	97

Both scenarios (1) and (2) refer to a study population of 1000 with a prevalence of disease of 5%, using the same test with a sensitivity of 50% and a specificity of 95% .

1.3.6 Results

Point estimates from studies in Table 3 indicate that HPV testing at the manufacturer's recommended threshold of positivity of 1 pg/ml (equivalent to 5000 viral copies) may be more sensitive than the Pap test (threshold of ASC-US) to detect CIN2 or worse lesions. The absolute difference in sensitivity ranged from 1% to 44%. This range is quite wide and can be attributed mostly to the wide range of sensitivity estimates for Pap testing across studies. However, differences in populations, screening tests and diagnostic procedures all contribute to the heterogeneity of the findings.

HPV testing, at the same threshold was consistently less specific (range 3% to 30%) than Pap testing. However, specificity estimates are difficult to interpret, since in most studies only women with one positive screening test underwent diagnostic procedures, and as such may not be representative of most test negative women in a general screening program.

Another interesting aspect of those study results is the fact that HPV testing estimates of sensitivity are more similar across studies (82%-100%) compared to those of Pap testing (40%-94%). This difference underlines the poor reproducibility of the Pap test and offers the possibility that HPV testing may also target this shortcoming.

1.4 Rationale and relevance

Cervical cancer screening is a topic of great concern in Canadian public health. Although, overall, cervical cancer rates in Canada compare favourably with the rest of the world [Parkin 2002], we have not been able to achieve optimal control throughout the country. Indeed, incidence rates in the Maritime Provinces are among the highest in the Western developed world [CCS 2007]. The wide range of sensitivity levels across laboratories, may explain why the Pap smear has not contributed evenly to cervical cancer control across Canada. Also, there is a growing concern that cervical cancer rates are increasing among more recent birth cohorts that are entering the age when cervical cancer begins to peak, even in settings with proper Pap screening in place [Anttilla 1999, Liu 2001].

In an era of diminishing health care funding, it is imperative that we look for potentially more efficient cancer prevention strategies. Primary prevention through anti-HPV vaccination will most probably have an impact on cervical cancer control. However, there are several reasons why it will not obviate the need for screening, at least not anytime soon. VLP vaccines are prophylactic, with limited benefit for women who are, or have been, already infected by oncogenic HPVs included in the vaccine [Koutsky 2006, FUTURE 2007, Saslow 2007, Sawaya 2007]. Given that almost 50% of Canadian teenage girls report sexual activity by the age of 16 [Maticka-Tyndale 2001], and given that 40-50% of women acquire at least one HPV infection within 4 years of initiating sexual activities [Collins 2002, Winer 2006], it is possible that vaccination will have a minimal impact on the risk of cervical cancer for women who are today in their mid-twenties or older. These women will require regular screening for decades to come. Also, the first

generation of vaccines available on the market target only 2 of the 15-20 HPV types associated with cancer. Even if young women are successfully vaccinated before infection with an oncogenic HPV, continued screening is recommended since they will only be protected against a fraction of cancer causing viruses [Saslow 2007, NACI 2007].

The previous sections have made the point that screening can play an important role in cervical cancer control. Indeed, without screening, cervical cancer would be a very important cause of morbidity and mortality; the pre-clinical phase of the disease is well characterized and we know when it is best to intervene; we have access to diagnostic and therapeutic measures, which, combined, lead to an improved outcome.

Given these conditions, we would expect a near eradication of cervical cancer. However, 2 weak links in the screening chain of events may explain why cervical cancer is expected to claim close to 400 lives in Canada yearly, and be responsible for morbid therapeutic procedures in 1350 women diagnosed with invasive cancer [CCS 2007]. Firstly, some women do not get screened, and efforts should target groups known to be under screened. However, the fact that half of cervical cancer cases are diagnosed in women with adequate screening is a worrisome perspective [Stuart 1997, Sadler 2004a, Leyden 2005, Spence 2007]. The low sensitivity of the Pap smear explains why women who are regularly screened are still diagnosed with cancer [Shaw 2002, Cuzick 2006a]. Any improvement in cervical cancer control will thus have to address the limitations of the screening test currently used. Moreover, the low sensitivity of the Pap smear leads to inefficient allocation of resources. Indeed, to offset the frequency of false negative results, some professional organizations recommend that women undergo over 50 rounds

of screening (yearly, from age 18 to 69) for cervical cancer [Parboosingh 1999, Anttilla 2004, Miller 2002, Smith 2006]. In this situation, the advantage of the high specificity of the Pap smear is lost, as many women will have at least one false positive result in their lifetime leading to unnecessary diagnostic investigation. Resources spent on yearly screening and ensuing follow-up could be better utilized to reach underserved groups.

As described in the previous sections, HPV testing with HC2 is a well suited candidate to replace Pap testing in cervical cancer screening: it appears to be a more sensitive test, has an acceptable specificity, and it is reproducible. It can be automated for high throughput leading to cost savings. However, characteristics of the initial studies on this topic made it difficult to translate their results into gains that could be expected if HPV testing was used to screen women in Canada.

The Canadian Task Force on Periodic Health Examination last reviewed the matter in 1995 and concluded at that moment that it was still premature to adopt large scale HPV testing in our Canada [Johnson 1995]. However, the Task Force established as one of the key research priorities the "assessment of efficacy and cost-effectiveness of screening for HPV infection." Given the complexities of undertaking the trial that would answer this pressing question, this information is still not available in Canada, more than 10 years after it was deemed a priority. The following research project addresses the efficacy portion of this question directly.

2 OBJECTIVE OF THE THESIS

Primary objective:

Compare, using an RCT study design, the performance (sensitivity, specificity, positive and negative predictive values) of HPV testing vs. Pap cytology in the detection of cervical cancers and their high-grade precursors among women aged 30-69 years who present for routine cervical cancer screening in Montreal and in St. John's

Secondary objectives:

Explore the impact of (1) sampling order, (2) test thresholds and combinations, and (3) patient and laboratory characteristics on test performance.

3 METHODOLOGY

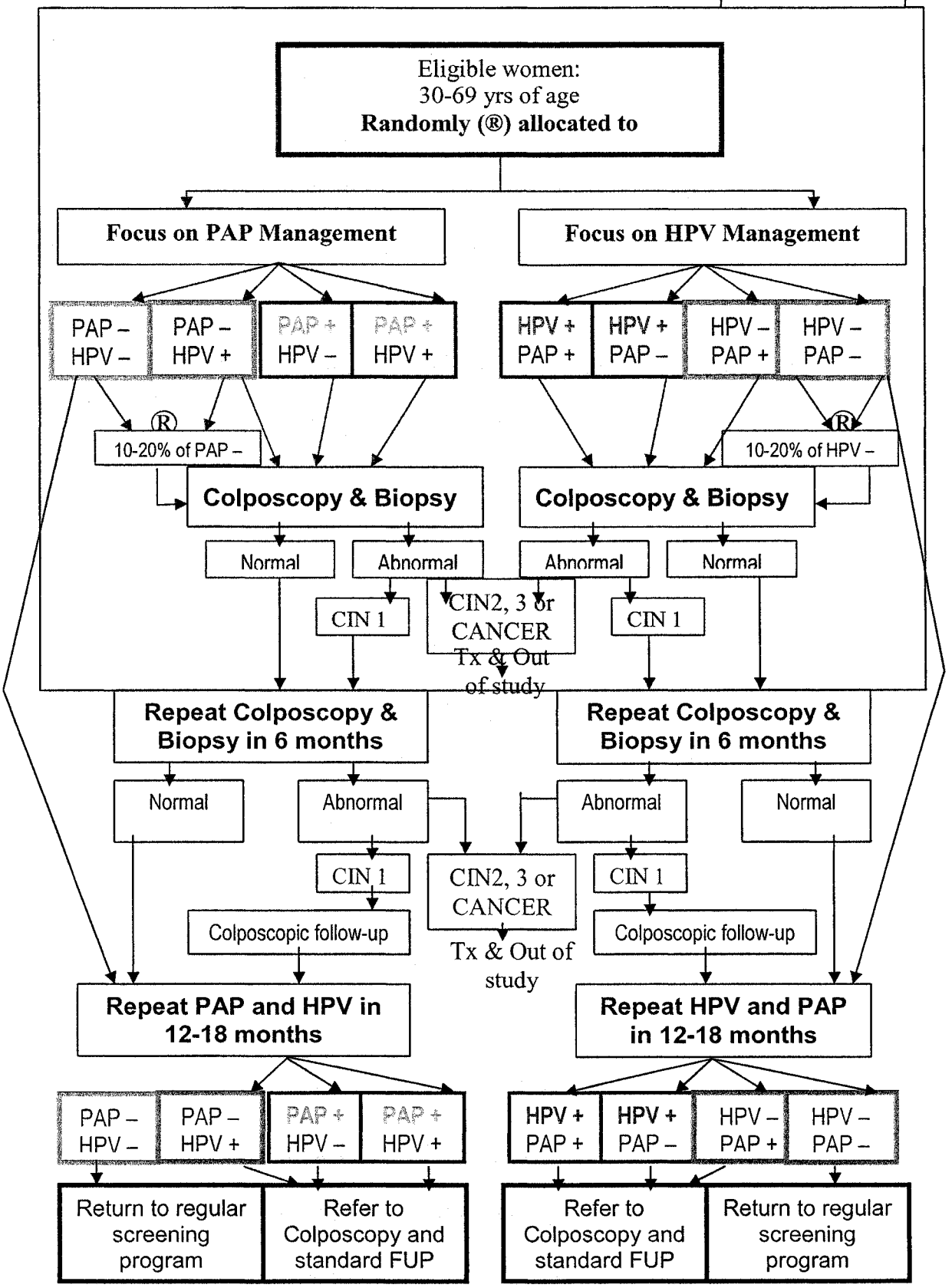
3.1 Design

The Canadian Cervical Cancer Screening trial or CCCaST, is an RCT which was designed to compare Pap and HPV testing as standalone tests for cervical cancer screening (International Standard Randomized Controlled Trial Number 57612064). CCCaST includes a cross-sectional analysis and a follow-up component. Figure 7 gives a schematic outline of the trial. This thesis work focuses on the cross-sectional component, which is described below.

Participants were randomized 1:1 to one of two arms, designated the "focus on Pap" screening arm (Pap arm for short) or the "focus on HPV" screening arm (HPV arm for short). To compare the efficacy of HPV testing to that of Pap testing, the simplest design would have been to randomize women to one of 2 screening strategies: Pap or HPV testing. However, at the onset of CCCaST, there was insufficient evidence regarding the efficacy of HPV testing as a stand-alone screening test to withhold Pap cytology from women in the trial. For this reason, we included both tests in each arm, but **randomized** the order in which the test samples were collected. In the "focus on Pap" screening arm, women received a Pap test (the index test) followed by an HPV test (the secondary test). In the "focus on HPV" screening arm, women received an HPV (the index test) followed by a Pap test (the secondary test). This design provided, at the analysis stage, the possibility to assess the performance of the two tests as if they had been done alone, while giving all women in the trial access to the established standard in cervical cancer screening: the Pap test. It also enabled us to investigate the performance of the 2 tests when used in combination and evaluate any biasing effects due to the test sampling order.

Figure 7. Outline of study design for the CCCaST trial

Focus of thesis



Women with an abnormal Pap test (the types of cytological abnormalities that were considered as a positive Pap test are detailed below) or a positive HPV test at enrolment underwent a colposcopic examination and biopsies. If cancer, CIN2 or CIN3 was found they were managed as per standard practice. A random sample of women testing negative with the index test also underwent colposcopy, to allow for correction of verification bias.

CCCaST was approved by the ethical review boards of each participating hospital and clinic, and of McGill and Memorial Universities (Appendix B).

3.2 Study Population

The study population was comprised of women 30 to 69 years of age, from Montreal and surrounding municipalities (province of Quebec) and St. John's (province of Newfoundland), enrolled through 30 selected medical practices. The study was limited to women older than 30, since in sexually active women younger than 30 years old, transient HPV infections are common, which gives HPV testing an unacceptably low specificity in identifying cervical cancer precursors [Cuzick 1995, Cuzick 1999a, Schiffman, 2000, Ratnam 2000]. The upper age limit (69 years) was set in accordance with current Canadian guidelines [Miller 1991].

Cytology laboratories in the study regions provided us with a list of physicians requesting Pap tests. This allowed us to identify and invite physicians from medical practices that were active in cervical cancer screening. This list included university-affiliated physicians and those in private practice, from family medicine and gynaecology practices of

different sizes that focus on primary care. This diverse base for recruitment ensured that the screening test efficacy measured in the study corresponded to a broad cross-section of providers.

The target population included women aged 30-69 years consulting in the participating medical practices. We excluded from the study women who were: (i) attending a colposcopy clinic for evaluation, treatment, or follow-up of a cervical lesion, (ii) without a cervix, (iii) pregnant, (iv) with a previous history of invasive cervical cancer, and (v) unable to provide informed consent. To maintain our focus on women undergoing routine screening, we also excluded women who had received a Pap test in the 12 months prior to enrolment. Indeed, women receiving a second Pap smear within a 12 month period are less likely to be undergoing routine screening, and more likely to be consulting for symptoms or for a previous Pap abnormality.

3.3 Recruitment procedures

Whenever CCCaST study group members or research assistants were available, potentially eligible women were given an information brochure and a self-administered enrolment card to determine eligibility (Appendix C). They were trained to clarify the study procedures and obtain written informed consent (Appendix B) from interested eligible women.

3.4 Baseline information

Participants also completed a self-administered questionnaire that elicited data on baseline demographics, risk factors for HPV infection and cervical cancer and information to be used in a subsequent in-depth cost analysis (Appendix C). Previous research experience with the same population in Newfoundland had shown that the inclusion of certain reproductive health questions hindered recruitment, and for this reason a shorter questionnaires was developed for St John's.

3.5 Randomization

We used a computer-generated block randomization algorithm with block sizes that varied randomly. Randomization was stratified by practice 1:1 to one of the two screening arms. Randomization was carried out at the study coordination centre, and opaque and sealed envelopes were left in each practice. After an eligible woman had agreed to participate, physicians opened consecutive envelopes in order to determine arm allocation. The following study characteristics helped ensure that physicians adhered to the randomized group allocation: (i) all women received both screening tests, (ii) follow-up was the same in the two arms, (iii) physicians did not favour any one particular order of sampling. The "envelope" randomization strategy was chosen after it became clear, through consultation with collaborating community physicians, that calling a central number for group allocation would not be feasible in the context of brief medical consultations such as those of "routine" visits for healthy women.

3.6 Blinding

Patients were blinded to arm allocation. Cytotechnicians and cytopathologists evaluating the Pap smears were not aware of inclusion of women into our study to ensure that the study samples would be treated no differently than other routine tests. They had no access to HPV results and HPV testing sites were unaware of other results. Colposcopists and pathologists evaluating the biopsy specimens were blinded to initial screening test results and arm allocation.

3.7 Study procedures: screening tests

3.7.1 Pap tests

The CCCaST study was designed to provide an evaluation of screening tests as they are typically performed in the community. For this reason, the procedures for Pap test collection, smear preparation, and processing were not standardized across centres. Conventional Pap tests, where a scraped cervical sample is smeared on a glass slide, fixed and shipped for staining and microscopic lecture, were in use in both study sites. Community cytotechnicians and laboratory pathologists, unaware of inclusion into a study and of HPV results, reported the Pap results using prevailing nomenclature and forms. At the onset of the study in October 2002, most cytology laboratories were using the Bethesda 2001 terminology, with only a few Montreal laboratories still using the Bethesda 1991 nomenclature. All such test results were re-classified according to

Bethesda 2001 terminology [Solomon 2002, summarized in Table 2] by the study coordinator].

3.7.2 HPV tests

The Hybrid Capture2 test (HC2) (Digene, Inc., Gaithersburg, Maryland, USA) was selected for HPV testing for this study. It is the test that has been most extensively evaluated in clinical practice and was the only HPV test approved by the Federal Drug Administration (FDA) of the United States. The test procedure is well standardized and can be carried out in most clinical microbiology laboratories.

Specimens were collected with the Digene cervical sampler kit which contains a specially designed cervical cytobrush and the specimen transport medium (STM). Physicians were instructed on how to collect the cervical specimen for HC2 testing (inserting the cytobrush in the cervical canal, turning it 3 times, then breaking off the brush in the transport vial), since most of them had never used this system. Specimens were stored at room temperature and shipped every two weeks to the two laboratories for analysis (Dr. Coutlée's at Hôpital Notre-Dame du Centre Hospitalier de l'Université de Montréal and Dr. Ratnam's at the Newfoundland Public Health Laboratory in St. John's).

The HC2 assays were performed according to the manufacturer's recommendations. Cervical cells in STM were denatured and 75µl of processed sample were hybridized with 25 µl of probe B mixture (containing RNA probes for high-risk HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68). After hybridization, specific DNA-RNA hybrids were

captured and detected onto wells of a microtiter plate using anti-DNA-RNA specific antibodies to capture DNA-RNA hybrids and a second set of anti-DNA-RNA antibodies conjugated with alkaline phosphatase to detect captured hybrids. The assays were completed with the addition of a dioxetane-based chemiluminescent substrate. As the substrate is cleaved by the bound alkaline phosphatase, light is emitted that is measured in a luminometer in relative light units (RLU). The intensity of the light emitted is proportional to the amount of HPV in the specimen. Specimens were considered positive for HPV if the ratio of RLUs of the specimen to the mean RLUs of triplicates of a positive control is equal to or greater than 1, as recommended by the manufacturer. This is equivalent to 1 pg HPV per ml or 5000 copies of HPV genome per test. Technicians and investigators at these 2 laboratories were unaware of cytology results when reporting HPV results.

We also performed a quality control assessment of the initial HPV test results provided by the two laboratories for study participants. The performance of the two laboratories was compared by having the two centres each exchange 80 samples, for a total of 160 pairs of results. Retesting of specimens from one centre was performed by the laboratory from the other centre and the results correlated. The specimen lists were prepared by the data management centre in Montreal and contained an approximately equal number of originally positive and negative results, chosen randomly. Retesting by each laboratory was done blindly. The 160 pairs of RLU results were analyzed by linear regression to assess the correlation between results (and departure from equivalency of results) on the basis of log-transformed data. The Kappa statistic was used to ascertain the agreement

between the two laboratories using RLUs recoded as negative, low positive (1<RLU<10), and high positive (RLU≥10).

3.8 Diagnostic assessment

Participants were contacted for colposcopy as results became available at the study coordinating centre. They were advised that, per protocol, they needed colposcopic evaluation, and that at the end of that examination the colposcopist would give them their diagnostic impression. Participants could also contact the study coordinating centre after the colposcopy to enquire about their screening test results. However, in the case that the participant refused colposcopy despite obtaining all the standard information, the study coordinating staff were instructed to inform patients of abnormal screening test results, if present, and explain the significance of those results.

Diagnosis was obtained through colposcopy-guided biopsies, the accepted standard for diagnosing cervical precancerous lesions [Wright 2002, Massad 2003]. Women with the following test results were contacted by CCCaST personnel and given an appointment for colposcopy at one of the participating colposcopy centres: those with a results of ASC-US, ASC-H, AGC, LSIL, HSIL, AIS or cancer on the Pap test (cytology results terminology is detailed in Table 2); those with a positive HPV test as described above; or those having been randomly selected for colposcopy despite the fact that both screening test results were negative.

Colposcopies took place at pre-selected centres, following a standardized protocol to reduce misclassification of disease status and verification bias. The protocol included (i) exo-cervical biopsies of all abnormal-appearing cervical regions, (ii) at least one exocervical biopsy aiming for aceto-white regions of cervices with no abnormalities (to reduce the risk that CIN giving an impression of metaplasia would not be biopsied and missed), and (iii) endocervical sampling.

Additional excision diagnostic procedures were carried out in cases of significant discrepancies between cytology and histology, or to confirm diagnosis in cases of glandular lesions or micro-invasive disease, as indicated by current standards of practice [Wright 2003a]. Most high grade lesions were treated by LEEP. When ablative therapeutic procedures (cryotherapy or laser fulguration) were selected, confirmatory biopsies were performed at the treatment visit.

3.9 Case definitions

High grade CIN (CIN2 or worse) is the accepted surrogate endpoint for cervical screening and an actionable finding for clinical management [Wright 2003]. Identification and treatment of precursor lesions enables screening to reduce not only mortality, but also incidence. Although the ultimate goal of screening is to reduce mortality from cervical cancer, trials with such an end point would need to last at least a decade, preferably two. Given the logistic and costs associated with such a trial, it is justified to previously proceed with controlled trials with intermediate endpoints (such as high grade precursor identification) before embarking on a project focusing on reduction of mortality [Greenwald 1985].

We used two different definitions for case status: a conservative or a liberal interpretation. Included in the liberal case definition were all histologically-confirmed CIN of grades 2/3, AIS or cervical cancers based on any of the histology specimens. Because colposcopically directed punch biopsies have been shown to sometimes overestimate lesion severity [Costa 2003] we have also used a conservative case definition, which included only the above that were confirmed in the excision LEEP specimen or in the confirmatory biopsy in the case of ablative treatment.

Pathologists at each hospital where the colposcopy was done made the histological diagnosis for all biopsy specimens following standard histological criteria and without knowledge of group allocation or the test result(s) that elicited the diagnostic work-up. The integration of all colposcopy results (colposcopic impression, cytology, ecto and endocervical histology, excision specimens) was done at the study coordinating centre, without knowledge of group allocation or screening tests results.

3.10 Sample size calculation

We used sensitivity as the primary parameter to guide sample size calculations. We wanted the trial to be powered to detect, at least, a difference in sensitivity of 20%. Given the costs and complexities of switching from cellular to viral tests in cervical cancer screening, a difference of less than 20% would probably not be sufficient to warrant a change in clinical practice. The unweighted average difference in sensitivity between Pap and HPV tests across studies free of verification bias (Table 3) [Ratnam 2000, Belinson 2001, Blumenthal 2001] is 29%. Thus finding a difference of 20% was deemed plausible.

The best estimate of Pap sensitivity, based on the meta-analyses of Nanda [2000], is 50%. This is slightly higher to the 40% reported by Ratnam [2000] in a population similar to the one targeted by this trial. Cuzick [1999a], Ratnam [2000], and Clavel [2001] reported a frequency of histologically verified high grade (or worse) of 1.4%-1.6%. These estimates are the most pertinent since they are derived from populations with similar prior access to screening.

A conservative calculation using a 20% difference in sensitivity at a 60% baseline sensitivity yields N=80 high grade lesions (or worse) as outcomes in each arm. Assuming a prevalence of 1.6%, this translates into N=4967 participants to be admitted in each arm to enable 80% power to resolve between sensitivity estimates of Pap and HPV. In order to accommodate losses to follow-up (up to 15%) the target sample size was set at 6000 subjects per arm, for a total of 12,000 women. Since specificity estimates are based on women judged free of lesions, this sample size was deemed adequate for the comparison of the latter index between groups.

In order to correct for verification bias we needed to invite participants who had a negative result in the index test in each arm to attend colposcopy. The fact that there had been almost no diagnostic evaluation of screen-negative women reported in the literature made it impossible to calculate the appropriate sample size based on firm data. We made a pragmatic decision guided by the fact that on one hand we wanted a sufficient number of diagnostic evaluations to be confident in our estimation of prevalence in those population, but on the other hand we did not want to deter recruitment by having a too high proportion of women undergo colposcopy and biopsy, and so set our target to have

colposcopic evaluation of 10% of the study sample that tested negative on the index tests We expected up to 30-50% of women to refuse colposcopy despite counselling both at enrolment and when contacted for colposcopy. For this reason, 20% of index test negative women were invited for colposcopy in Montreal but limited colposcopy resources precluded the same approach for St. John's. Thus the final group of test negative women to be invited for colposcopy was comprised of a clinic-stratified, randomly selected 10% of index test negative participants in St. John's and 20% in Montreal. An important but tenable assumption for the correction approach to be valid is that refusal to undergo colposcopy is not linked to risk of outcome. This assumption was verified by comparing baseline demographic and risk factors between women who accepted and those who refused random colposcopy (Tables 11 and 12).

3.11 Analysis

3.11.1 Descriptive data

Where appropriate, possible differences in categorical data between groups were investigated using Fisher's and Chi-square tests, and differences in the medians of continuous data with the Kruskal-Wallis test. All tests were 2-sided.

3.11.2 Main analysis

Although the ability to identify an increase in sensitivity with HPV testing relative to Pap cytology was the key criterion for the sample size calculation, the main analysis consists of the comparison of four screening indices (sensitivity, specificity, and positive and

negative predictive values) of the 2 screening tests, Pap and HPV. The calculation of the four indices was conducted independently for each study arm, respecting original arm allocation (i.e., an "intention to treat" approach), including all women with an index screening test result for each arm (98.5% of all women in the Pap arm and 99.7% of those in the HPV arm). Three computation strategies are reported: (1) the **crude** estimates include only women who underwent diagnostic assessment (colposcopy). 2 x 2 tables compiling the joint results of testing procedures and disease ascertainment in each arm were constructed, which allowed the computation of the four indices with their respective asymptotic 95% confidence intervals based on the binomial distribution; (2) the **uncorrected** estimates include all women with an index test result in each arm and assume that all participants who did not undergo diagnostic evaluation were free of disease. Again, 2 x 2 tables compiling the joint results of screening and diagnostic results were constructed and asymptotic 95% confidence intervals based on the binomial distribution calculated. Although biased, these first two evaluations are reported because they were used in previously published trials, and we wished to compare our results to those previously published. Also, it gave us the possibility to investigate the extent of the bias; (3) the third set of screening indices, termed **corrected** estimates, was corrected for verification bias and represents the most accurate estimation of the performance of the 2 screening tests, assuming that disease ascertainment among those who were test negative was free of error.

As detailed in section 1.3.5, verification bias is an important limitation of many screening studies. Because we could not perform diagnostic evaluation on all study participants (for ethical, costs and resource utilization considerations), we opted to perform disease

verification in a random sample of screen negative participants and use this information to calculate the likely number of cases that would have been found, if all screen negative participants had been fully investigated. This allowed the estimation of the number of additional cases that would have been found if all participants had received the diagnostic evaluation. The key assumption of this strategy is that the disease prevalence in each stratum is assumed to be independent of whether women underwent colposcopy or not. This strategy has previously been described in detail and used in other cervical cancer screening studies [Ratnam 2000, Kulasingam 2002].

To compute the **corrected** estimates, data were divided in strata defined by combined Pap and HPV results. Stratum-specific probabilities were then applied to the remainder of women who had not undergone colposcopy, which permitted estimating the number of cases that would have been found if all study participants had undergone histological verification. Corrected sensitivity and specificity estimates were then calculated and 95% CIs computed by the method described by Zhou [Zhou 1998].

3.11.3 Exploratory analysis

3.11.3.1 Interference of sampling order

We then investigated if the order of specimen collection for the tests had an effect on their performance. To do this, we compared the overall proportion of positive tests, crude sensitivity and specificity estimates, proportion of unsatisfactory and equivocal smears for Pap testing and distribution of RLU levels for HPV testing. These characteristics are

indicative of the adequacy of cervical sampling. Differences in categorical data between study arms and/or centres were investigated using Fisher's and Chi-square tests, and differences in the medians of continuous data with the Kruskal-Wallis test. All tests were 2-sided. 95% CIs for the difference between proportions were also computed, to assess statistical significance as well as to help judge clinically relevant differences.

3.11.3.2 Estimates of screening performance, varying thresholds and combinations

Corrected indices were computed, according to the technique described in section 3.8.2, varying the positivity threshold for Pap and HPV. Different testing sequences and combinations were also explored.

Analyses were performed with SAS 9.1 (SAS Institute Inc., Cary, NC, USA) statistical software.

3.11.3.3 Impact of patient and laboratory characteristics on test performance

Model: To identify spectrum effect [Ransohoff 1978], that is patient and laboratory characteristics which may influence the performance (sensitivity and specificity) of Pap and/or HPV testing, logistic regression modeling (LRM) was used. Stratification would have been impractical given the relatively small number of cases, as the impact of only a limited number of parameters and combinations could have been investigated. On the other hand, LRM makes use of all available data, and provides smoothed estimates of

sensitivity and specificity. Sensitivity and specificity were chosen as key parameters to explore because they are thought to be "intrinsically" related to test performance, compared to predictive values which are influenced by prevalence [Zhou 2002].

To model sensitivity and specificity, the approach described by Coughlin was used [Coughlin 1992]. Namely, the dependent variable of the model was the dichotomous screening test result. The case status (dichotomous diagnostic test variable) was included as an independent variable in all models. Other variables (defined in section below) that could influence test performance were then entered as independent variables also, as follows:

$$\text{Ln } \{P(T=1)/[1-P(T=1)]\} = \alpha + \beta_1 X_1 + \Sigma \beta i Xi$$

Where α denotes the estimated intercept and β_i's denote the estimated coefficients for the array of covariates included in the model, factored as per the indicator variables X_i's. X_1 is by definition always included in any model and refers to the diagnostic variable (case status).

Once the final model is selected, sensitivity, specificity and their 95% CIs can be computed for the different subgroups according to the following formulas:

$$\text{Sensitivity} = \frac{1}{1 + \exp\left[-\left(\alpha + \sum_{K=1}^{K} \beta_1 K_1 + \beta_i K_i\right)\right]} \qquad \text{Where } X_1 = 1.$$

Sensitivity= $1-\left\{\frac{1}{1+\exp[-(\alpha+\sum_{K=1}^{K}\beta_1 K_1+\beta_i K_i)]}\right\}$ Where $X_1 = 0$.

95% confidence interval for estimated sensitivity at covariate level X_2^*:

$$\frac{1}{1+\exp[-(\alpha+\beta_1+\beta_2 X_2^*) +/- 1.96\sqrt{\sigma}\,2\,]}$$

[Coughlin 1992].

Dataset: To investigate the impact of variables on test performance, we used the full CCCaST dataset (irrespective of arm allocation). Models were done separately for Pap and HPV. In the Pap model, all participants with a Pap test result were initially included (9 991). Participants with one or two positive screening test result(s) who did not undergo diagnostic evaluation were then excluded from the dataset (23 excluded, final dataset n=9968). Participants with 2 negative results on both screening tests were assumed to be free of disease (conservative case definition). In the HPV model, all participants with an HPV test result were initially included (10120). Participants with one or two positive screening test result(s) who did not undergo diagnostic evaluation were then excluded from the dataset. (50 excluded, final dataset n=10070). Participants with 2 negative results on both screening tests were assumed to be free of disease (conservative case definition).

Variables: Variables to be included in the model were selected because of documented or potential impact on test performance. Various laboratory characteristics and factors which influence Pap update have been identified [Peters 1988, Brinton 1994, Jepson 2001, Breen 2005, Forbes 2007, Tacken 2007]. However, we could not identify literature on patient characteristics which influence Pap nor HPV performance. However, factors that influence the natural history of HPV infection are well described: age, markers of sexual activity, smoking and use of hormonal contraception [Ho 1998a, Hildesheim 2001, Castellsague 2003]. Any characteristic which increases HPV incidence and/or prevalence, but has a limited impact on risk of disease was of particular interest.

The dependent variable of each model was the dichotomous screening test result. Potential independent variables were (1) presence or absence of disease according to the conservative case definition (we chose the conservative case definition because we believe it is more accurate-however all analysis were rerun with the more liberal case definition, and results were similar); (2) age; (3) present smoking status; (4) marital status; (5) education; (6) contraception; and (7) laboratory which analyzed the test. The association with age was not linear, and for this reason age was introduced in the model as a categorical variable (30s, 40, 50s, and 60s). Complex approaches to missing data, such as multiple imputations, have been suggested in the case when "a large proportion of data are missing" [Greenland 1995]. Given that less than 2% of data were missing and that over 97% of subjects had complete data information, a "missing" category was created for each of the independent variables, except (1). All categorical variables were represented in the models by their equivalent set of k – 1 "dummy" indicator variables, where k denotes the number of categories.

Table 5. Initial variable categories considered for LRM

Variable	Type	Categories	Reference category
Screening test result	Dichotomous	Positive	
		Negative	*
Diagnostic test result	Dichotomous	Present	
		Absent	*
Age	Categorical	30-39	*
		40-49	
		50-59	
		60-69	
Smoking	Dichotomous	Current smoker	
		Current non smoker	*
Marital status	Categorical	Married/living in union	*
		Single	
		Separated, divorced/widowed	
Schooling	Categorical	Primary school	*
		High school	
		College	
		University	
Contraception	Categorical	Sterilization	
		Hormonal	
		IUD	
		Barrier	
		Natural	
		None	*
Pap laboratory	Categorical	Laboratory 1	*
		Laboratory 2	
		Laboratory 3	
		Laboratory 4	
		Laboratory 5	
		Laboratory 6	
		Laboratory 7	
		Laboratory 8	
HPV laboratory	Dichotomous	Laboratory 1	*
		Laboratory 2	

Model building: Backward selection based Akaike's Information Criteria (AIC) [Akaike 1977] used to select the final models, one model for Pap and one model for HPV. Forward AIC selection and stepwise selection based on the Wald test were also run (with p=0.1 for entry and p=0.15 for removal) to check for consistency, and similar results were obtained. Once the final model was selected, individual coefficients of categorical variables were verified to check if certain categories could be grouped. When combination of categories made sense clinically, and coefficients were similar, regrouping was done in order to obtain the most parsimonious model. Backward AIC selection was retried after recoding to confirm that the same variables would be selected for the final model.

No interaction was defined a priory as being essential. All possible interaction terms between independent variables were entered separately as potential variables for model selection. None reached statistical significance and so none was retained for the final model.

The coefficient of the variable "presence or absence of disease" was compared in the initial "full" model and in the final model to ensure that the removal of variables had not created confounding. The coefficient varied by less than 10%, and for this reason, no variable was forced back in the model.

The same method was used to select the final model for the analysis restricted to Montreal participants. Only one variable was added: "number of new sexual partner in the last year". Because of non-linearity, the variable could not be included as a continuous

variable. Categories were created as follows, (1) none, (2) one, (3) two, and (4) three or more. The referent category was "none". This new variable was selected in the final model, and the coefficient of all categories was similar. For this reason, the variable was recoded as "none" vs. "one or more".

Stata 9.2 (Stata Corp., College Station, Texas, USA) was used for LRM analyses.

4 RESULTS

4.1 Participants characteristics

Between September 26th, 2002 and February 3rd, 2005, 14,953 women were assessed for eligibility in the CCCaST trial (Figure 8). A total of 4782 candidate participants were ineligible or refused: 184 had a recent Pap smear, 440 were outside the eligible age range, 874 had been followed in a colposcopy clinic in the last 2 years, 691 had had a total hysterectomy, 90 were pregnant, 86 were unable to provide informed consent, and 1267 refused (for some women there was more than one reason for exclusion). This left 10171 participants who were randomized to one of the two trial arms. After randomization, file review showed that 34 women had been randomized even though they were outside the target age range. Of these, 17 were less than a year short of being 30, and so were kept in the analysis. The other 17 were excluded, leaving a study population of 10154 participants. 5059 were randomized to the Pap arm and 5095 to HPV arm. As shown in Figure 8, over 98% of the participants received the allocated intervention as intended (99.2% in the Pap arm, 97.2% in the HPV arm). In both arms, most participants with an incorrect intervention actually received both tests, albeit in the wrong order.

In the Pap arm, 388 women needed colposcopy because of an abnormal screening test compared to 396 in the HPV arm. In the Pap and HPV arms, 706 and 664 participants selected from among those with negative screen results were invited to attend a colposcopic examination. Over 90% (713/784) of participants with at least one positive result and 7.0% (652/9370) of those with negative test results underwent colposcopy.

Figure 8. Diagram of study progress

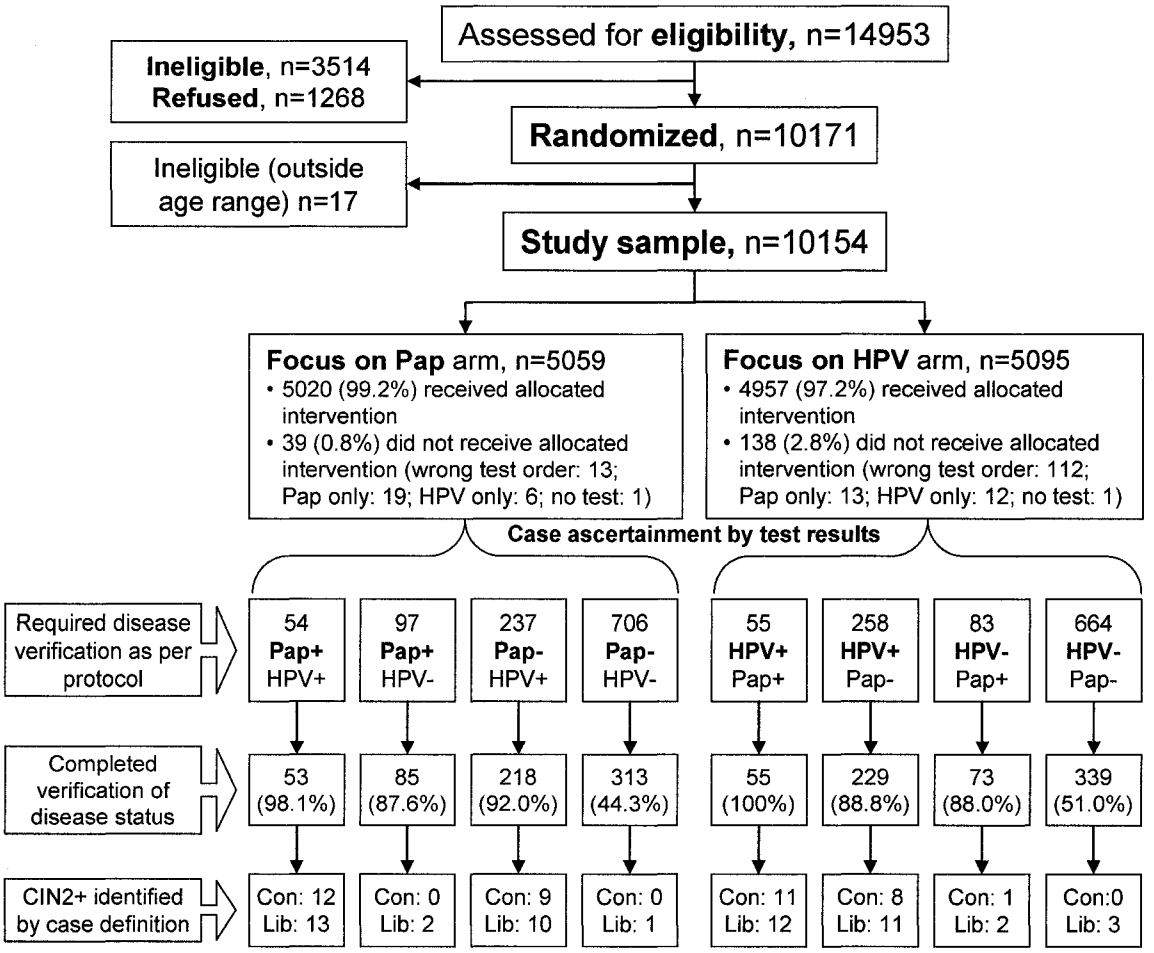


Figure Legend

Flow diagram of first screening round, cross-sectional component of CCCaST. Pap positive results (Pap+) defined as ≥ ASC; HPV positive (HPV+) defined as ≥ 1pg/ml; Con: conservative case definition; Lib: liberal case definition. Random selection for colposcopy for those with both negative tests as follows: 706 out of 4575 Pap-/HPV- participants and 664 out of 4600 HPV-/Pap- participants, for Pap and HPV arms, respectively. Participants who had only one test taken or who had a Pap test deemed unsatisfactory for interpretation are not included in the figure but were included in efficacy analyses. In the "focus on Pap" group, 95 participants had only one evaluable screening test: 19 had a negative Pap, 72 had a negative HPV test, and 4 had a positive HPV test. The latter 4 participants underwent colposcopy and none was found to have disease by either case definition. Among the former 91 participants with a negative screening test, 13 were randomly selected to undergo colposcopy; 3 complied and none was found to have disease by either definition. In the "focus on HPV group", 98 participants had only one evaluable screening test: 79 had a negative HPV test, 6 had a positive HPV test, 12 had a negative Pap test, and 1 had a positive Pap test. Among the 7 participants who had a positive test, 6 underwent colposcopy and 1 (HPV+) was found to have disease by the liberal case definition only. Among the 91 participants with a negative test result, 18 were randomly selected to undergo colposcopy; 10 complied and none was found to have disease by either definition. Both arms were balanced with respect with the required disease verification (p=0.5).

The latter group represents 47.6% (652/1370) of those invited following a random assignment to colposcopy.

Although no subgroup analysis was performed in regard to test performance (sensitivity, specificity, PPV, NPV) the following sections briefly explore the distributions of participants' characteristics and screening test results in regard to study arm but also study center, in order to provide a more detailed picture of the study population.

Table 5 presents selected participants' characteristics by arm and centre. Randomization produced comparable groups. However, differences between study centres are evident: in addition to the expected preponderance of French Canadians, Montreal participants were older than those in St. John's (medians: 44 vs. 42; $p<0.001$). Marital status differed between centres with more single women in Montreal than in St. John's (17.8% vs 8.6%; $p<0.001$). Smoking was more common in Montreal than in St. John's (23.1% vs 16.6%; $p<0.001$). Contraceptive choices differed between centres, with more participants in St. John's reporting no contraceptive use, and more women in Montreal reporting use of barrier methods. Over 96% of participants reported having had a Pap test in the past. Overall, 25% of participants reported having been told that their Pap test was abnormal at least once in the past. A longer questionnaire with items on reproductive health and sexual practices was administered only to Montreal participants. Of the latter, the median age at first intercourse was 18 and over 80% had been pregnant at least once. 28.0 % were menopausal and of those, close to half were taking some form of substitutive hormonal therapy.

Table 6. Distribution of selected characteristics of participants enrolled into the CCCaST study by arm and centre.

Characteristic	Categories	Montreal 4400		St. John's 5754	
		Focus on Pap arm (N=2191)	Focus on HPV arm (N=2209)	Focus on Pap arm (N=2868)	Focus on HPV arm (N=2886)
Age (years)[1]	30-39	732 (33.4)	746 (33.8)	1213 (42.3)	1215 (42.1)
	40-49	766 (35.0)	772 (34.9)	1028 (35.8)	992 (34.4)
	50-59	519 (23.7)	511 (23.1)	486 (16.9)	556 (19.3)
	60-69	174 (7.9)	180 (8.1)	141 (4.9)	123 (4.3)
Ethnicity[2]	French Canadian	1808 (82.5)	1831 (82.9)	17 (0.6)	18 (0.6)
	English Canadian	53 (2.4)	50 (2.3)	2792 (97.4)	2807 (97.3)
	Other	277 (12.6)	284 (12.9)	46 (1.6)	36 (1.2)
Schooling	Elementary school	243 (11.1)	221 (10.0)	287 (10.0)	289 (10.0)
	High school	991 (45.2)	808 (36.6)	629 (21.9)	656 (22.7)
	College	588 (26.8)	577 (26.1)	878 (30.6)	872 (30.2)
	University	835 (38.1)	871 (39.4)	1063 (37.1)	1056 (36.6)
Marital Status	Single	393 (17.9)	380 (17.2)	220 (7.7)	269 (9.3)
	Married/in union	1388 (63.4)	1442 (65.3)	2287 (79.7)	2324 (80.5)
	Separated/widowed	378 (17.3)	357 (16.2)	342 (11.9)	276 (9.6)
Contraception[3]	Sterilization[4]	316 (20.7)	343 (22.4)	590 (25.6)	597 (26.3)
	Hormonal[5]	367 (24.1)	318 (22.8)	500 (21.7)	483 (21.3)
	IUD[6]	77 (5.0)	91 (5.0)	33 (1.4)	34 (1.5)
	Barrier[7]	290 (19.0)	279 (18.2)	223 (7.8)	223 (9.8)
	Natural[8]	10 (0.7)	11 (0.7)	6 (0.3)	8 (0.4)
	None	465 (30.5)	488 (31.9)	952 (41.3)	926 (40.8)
Age at first intercourse	Median (IQR[9])	18 (16-20)	18 (17-20)	NA[10]	NA
	Never had intercourse	9 (0.4)	7 (0.3)	NA	NA
Lifetime no. of sexual partners	Median (IQR)	5 (2-10)	4 (2-10)	NA	NA
	Never had intercourse	9 (0.4)	7 (0.3)	NA	NA
Number of pregnancies	0	394 (18.0)	412 (18.7)	NA	NA
	1-2	1045 (47.7)	1068 (48.3)	NA	NA
	3-4	565 (25.8)	532 (24.1)	NA	NA
	> 4	119 (5.4)	137 (6.2)	NA	NA
Menopaused	Yes	610 (27.8)	620 (28.1)	NA	NA
	No	1381 (63.0)	1380 (62.5)	NA	NA
	Don't know	147 (6.7)	149 (6.7)	NA	NA
Menopaused, taking HRT [11]	Yes	281 (46.1)	307 (49.0)	NA	NA
	No	311 (51.0)	289 (46.6)	NA	NA
Current smokers	Yes	514 (23.5)	490 (22.2)	492 (17.2)	459 (15.9)
	No	1638 (74.8)	1687 (76.4)	2369 (82.6)	2415 (83.7)
Had Pap smear	Yes	2006 (91.6)	2020 (91.4)	2853 (99.5)	2869 (99.4)
	No	96 (4.4)	107 (4.8)	6 (0.2)	2 (0.1)
	Don't know	40 (1.8)	35 (1.6)	1 (0.0)	0 (0.0)

Characteristic	Categories	Montreal 4400		St. John's 5754	
		Focus on Pap arm (N=2191)	Focus on HPV arm (N=2209)	Focus on Pap arm (N=2868)	Focus on HPV arm (N=2886)
Self-report of ever having abnormal Pap	Yes	474 (21.6)	442 (20.0)	798 (27.8)	847 (29.3)
	No	1442 (65.8)	1471 (66.6)	1811 (63.1)	1760 (61.0)
	Don't know	49 (2.2)	71 (3.2)	135 (4.7)	142 (4.9)

Unless stated otherwise, numbers in parentheses represent percentages in all tables
[1] Percentages may not add up to 100% because of rounding and missing data
[2] Except for the age variable, there are some missing data, which were simply omitted in this table.
[3] Percentages for contraceptive categories based on number of women younger than 51 (median age at menopause for Canadian women); more than one method was used by some women
[4] Tubal ligation and vasectomy; [5] Oral and parenteral; [6] Intra-uterine device;[7] Male and female condom, diaphragm, cervical cap, contraceptive sponge;[8] Fertility awareness and coitus interruptus
[9] Inter-Quartile Range
[10]Not available, a shorter questionnaire was administered in St. John's
[11] The denominators for the percentages are women who reported being menopaused.

4.2 Screening test results

Table 7 summarizes screening test results. There were similar distributions of screening test result between study arms (Pap test dichotomous: 3.0% vs. 2.7% p value=0.5; HPV test dichotomous: 5.8% vs. 6.3%, p value=0.4) but results differed between centres. A higher number of participants had an abnormal HPV test result compared to abnormal Pap tests, 6.0% vs 2.9% overall ($p<0.001$). Overall, 2.9% of participants had an abnormal Pap test result (ASC-US or worse). The proportion of Pap tests deemed unsatisfactory to yield a cytological interpretation was low in both study arms (1.4% each arm). In both arms, the majority of abnormal cytology results were equivocal (ASC-US; 95/151 (62%) in the Pap arm and 92/139 (66%) in the HPV arm). Overall, over 97% of women with negative results on HPV testing had RLU readings below 0.75.

The proportion of positive Pap tests was 4.0% in Montreal but only 2.0% in St. John's ($p<0.001$). A higher proportion of Montreal women had positive HPV tests compared to participants from St. John's (7.7% vs 4.8%; $p<0.001$).

Table 7. Distribution of screening test results by arm and centre.

Group		Focus on Pap arm N=5059	Focus on HPV arm N=5095	Montreal N=4400	St. John's N=5754	Overall N=10154
Test or Test format	Result	N (%)	N (%)	N (%)	N (%)	N (%)
Pap test (dichotomous)	Negative	4831 (95.5)	4870 (95.6)	4155 (94.4)	5546 (96.4)	9701 (95.5)
	Positive[1]	151 (3.0)	139 (2.7)	175 (4.0)	115 (2.0)	290 (2.9)
	Unsatisfactory	70 (1.4)	73 (1.4)	55 (1.3)	88 (1.5)	143 (1.4)
	Not taken	7 (0.1)	13 (0.3)	15 (0.3)	5 (0.1)	20 (0.2)
Pap test (detail of categories)	Negative	4831 (95.5)	4870 (95.6)	4155 (94.4)	5546 (96.4)	9701 (95.5)
	ASC-US	95 (1.9)	92 (1.8)	118 (2.7)	69 (1.2)	187 (1.8)
	ASC-H	5 (0.1)	4 (0.1)	3 (0.1)	6 (0.1)	9 (0.1)
	AGC	24 (0.5)	22 (0.4)	27 (0.6)	19 (0.3)	46 (0.5)
	LSIL	17 (0.3)	15 (0.3)	18 (0.4)	14 (0.2)	32 (0.3)
	HSIL	9 (0.2)	6 (0.1)	8 (0.2)	7 (0.1)	15 (0.1)
	Sq Ca[2]	1 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	1 (0.0)
HPV test (dichotomous)	Negative	4744 (93.8)	4762 (93.5)	4032 (91.6)	5474 (95.1)	9506 (93.6)
	Positive[3]	295 (5.8)	319 (6.3)	337 (7.7)	277 (4.8)	614 (6.0)
	Not taken	20 (0.4)	14 (0.3)	31 (0.7)	3 (0.1)	34 (0.3)
HPV test (RLU results)	<0.75	4642 (91.8)	4643 (91.1)	3896 (88.6)	5389 (93.7)	9285 (91.4)
	0.75-0.99	102 (2.0)	119 (2.3)	136 (3.1)	85 (1.5)	221 (2.2)
	1.00-1.99	56 (1.1)	67 (1.3)	66 (1.5)	57 (1.0)	123 (1.2)
	2.00-3.99	41 (0.8)	45 (0.9)	44 (1.0)	42 (0.7)	86 (0.8)
	4.00-9.99	38 (0.8)	37(0.7)	44 (1.0)	31 (0.5)	75 (0.7)
	10-39.99	59(1.2)	62 (1.2)	66 (1.5)	55 (1.0)	121 (1.2)
	≥40	101(2.0)	108 (2.1)	117 (2.7)	92 (1.6)	209 (2.1)
Pap and HPV (dichotomous)	Both neg.	2146 (42.4)	2156 (42.3)	2653 (60.3)	2662 (46.3)	9175 (90.4)
	Only HPV pos	128 (2.5)	139 (2.7)	36 (0.8)	36 (0.6)	180 (1.8)
	Only Pap pos.	61 (1.2)	47 (0.9)	109 (2.5)	119 (2.1)	495 (4.9)
	Both pos.	35 (0.7)	31 (0.6)	19 (0.4)	24 (0.4)	109 (1.1)

[1] ≥ASC
[2] Squamous Carcinoma
[3] RLU ≥ 1.0

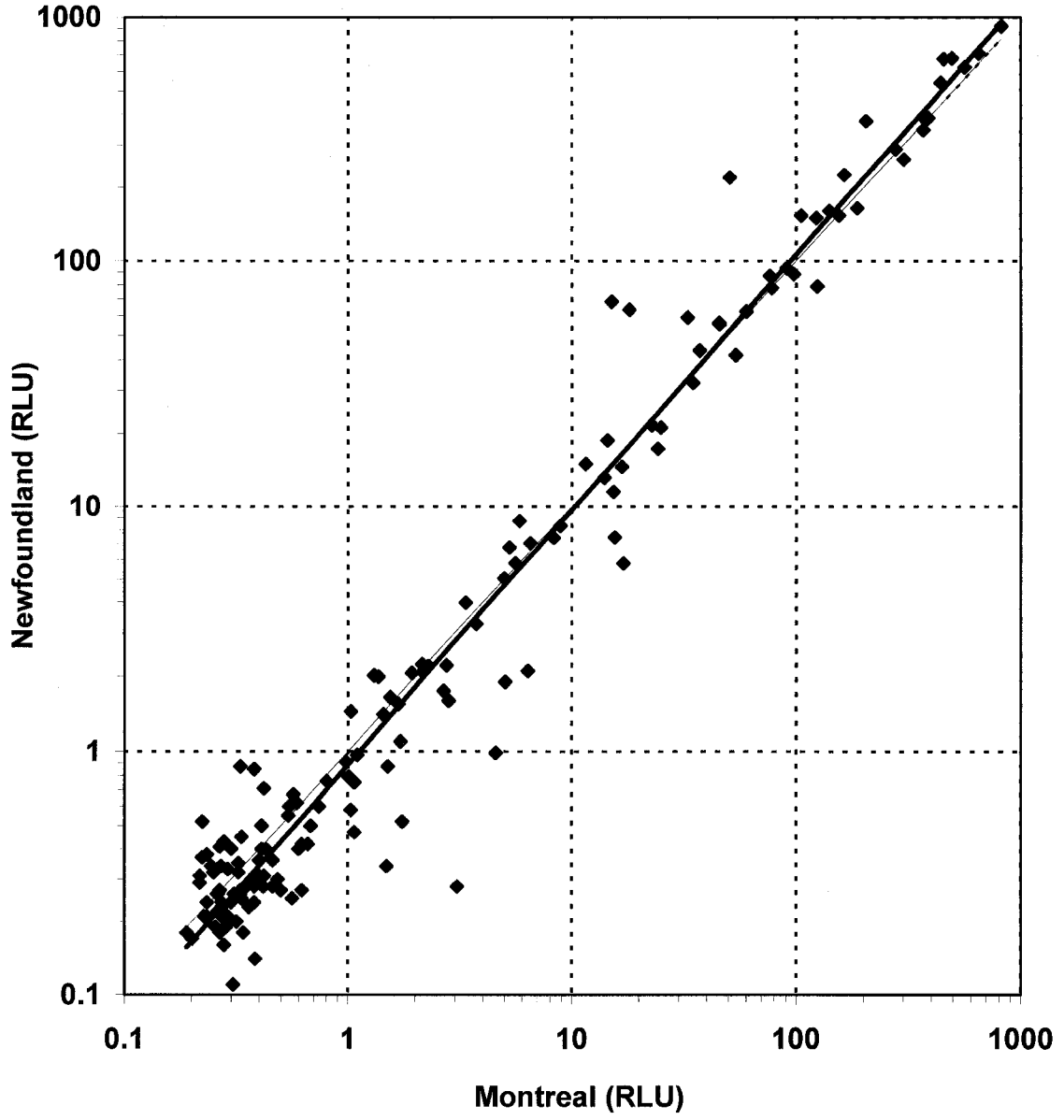
Table 8 explores the relationship between test results and age. The observed higher proportion of HPV positive women in Montreal could be explained through 2 hypotheses: a difference in age distributions (HPV prevalence is known to decrease with age) and/or technical laboratory considerations. Figure 9 illustrates that the differences in HPV positivity are real and not related to testing laboratory biases. There was a strong correlation between test results in one lab and retest by the other, with a near equivalency in RLU readings, as shown by the best fitting regression equation (Predicted Log St. John's RLU = -0.0542 + 1.0403 x observed Log Montreal RLU, $R^2 = 0.9671$). There was a high agreement between the two laboratories (Kappa = 0.878) when results were treated in 3 categories (negative, 1-10, > 10) based on recoding the RLUs.

Table 8. Proportion of positive screening tests by age group and centre

	Montreal			St. John's		
Age group	n	% Pap positive	% HPV positive	n	% Pap positive	% HPV positive
30-39 years	1478	4.5	12.6	2428	2.1	6.9
40-49 years	1538	4.5	5.8	2020	2.4	4.0
50-59 years	1030	2.9	4.7	1042	1.2	2.4
60-69 years	354	2.8	3.7	264	0.8	1.5

The possible influence of the difference in age distribution between centres was also evaluated as a possible explanation for the between-centres differences with respect to HPV positivity. However, a higher prevalence of HPV in Montreal could not be explained by differences in age.

Figure 9. Correlation between original and retest HPV results (RLU readings) for two samples of 80 specimens, one from each study centre.



Samples included approximately 40 positive and 40 negative results randomly chosen from each centre. Retesting of specimens from one centre was performed by the laboratory from the other centre. The thick line represents the regression model that best fits the data points. The thin line shows the expected equivalency between the two sets of results.

Indeed, the St. John's population was in fact younger (77.3% below 50 years of age compared to 68.6% in Montreal). It is interesting to note that the higher proportion of abnormal test results (both Pap and HPV) in Montreal was observed in each age group. In both centres, the proportion of positive Pap tests decreased with age (from 4.5% in participants in their 60s to 2.8% in participants in their 30s, in Montreal; and from 2.1% to 0.8% for the same age groups, in St. John's). However, HPV positivity decreased more pronouncedly with age in both centres (from 12.6% to 3.7% in Montreal, and from 6.9% to 1.5% in St. John's). Therefore, the difference in overall HPV prevalence between centres (7.7% vs. 4.8%) occurred not because of a difference in age structure, but despite this difference. In fact, when a combined population structure is used for the purposes of direct age standardization, the adjusted HPV positivity rates are 8.1% and 4.6%, for Montreal and St. John's, respectively.

We also explored the association between cytology results and HPV positivity (Table 9). Only 6.5% of women with normal Pap tests in Montreal and 4.1% in St. John's had positive results for HPV tests. HPV positivity increased dramatically in women with abnormal cytology results: 78.1% of women with LSIL cytological diagnoses had positive high-risk HPV tests, while over 93.3% of women with HSIL cytological diagnoses tested positive for high-risk HPV.

4.3 Diagnostic assessment: process and findings

As the protocol mandated, the following participants were referred for diagnostic assessment with colposcopy and biopsy: (1)those with a Pap test result $\geq$ ASC-US;

(2)those with an HPV results of ≥ 1mg HPV/ml; (3) those flagged randomly for diagnostic evaluation, despite negative test results on both screening tests.

Table 9. HPV positivity by cytological result and by centre[1]

Centre	Cytology results	Study groups				Total by Centre	
		Focus on Pap		Focus on HPV			
		n	% HPV positive	n	% HPV positive	n	% HPV positive
Montreal	Negative	2050	6.2	2077	6.7	4127	6.5
	ASC-US	65	26.2	52	34.6	117	29.9
	ASC-H	2	100.0	1	0.0	3	66.7
	AGC	15	26.7	12	25.0	27	25.9
	LSIL	9	77.8	9	66.7	18	72.2
	HSIL	4	100.0	4	100.0	8	100.0
	Squamous carcinoma	1	100.0	0	NA	1	100.0
	Unsatisfactory	22	4.6	33	6.06	55	5.4
St. John's	Negative	2762	4.0	2781	4.28	5543	4.1
	ASC-US	30	16.7	39	35.90	69	27.5
	ASC-H	3	33.3	3	33.33	6	33.3
	AGC	9	22.2	10	20.00	19	21.0
	LSIL	8	75.0	6	100.00	14	85.7
	HSIL	5	100.0	2	50.00	7	85.7
	Unsatisfactory	48	6.2	40	7.30	88	6.8

[1] This table includes data from the 10102 participants with both Pap and HPV tests taken

Table 10 provides key parameters of the diagnostic assessment process. Over 90% (723/795) of participants with at least one positive result underwent colposcopic evaluation, with similar proportions in the 2 study arms (90 vs. 92%; pvalue=0.39). 7.1% (665/9359) of those with negative test results underwent colposcopy (6.8 vs. 7.4; p value=0.21). The latter group represents 47.4% (665/1402) of those invited following a

Table 10. Diagnostic assessment process by study arm and screening test result

	Focus on Pap arm				
	Pap+/HPV+ N=54	Pap+/HPV- N=97	Pap-/HPV+ N=237	Pap-/HPV- N=4575	Missing Result(s) N=96
Colposcopy completed, N (%)	53 (98)	85 (88)	218 (92)	313 (7)	7 (7)
Histology available, N (%)	51 (96)	80 (94)	208 (95)	301 (96)	7 (100)
Time to colposcopy[1] (in days)	97	114	105	216	60
Number of cases (liberal)	13	2	12	1	0
Time to treatment[2] (in days)	47	108	62	153	NA
Number of cases (conservative)	12	0	11	0	0
	Focus on HPV arm				
	HPV+/Pap+ N=55	HPV-/Pap+ N=83	HPV+/Pap- N=258	HPV-/Pap- N=4600	Missing Result(s) N=99
Colposcopy completed, N (%)	55 (100)	73 (88)	229 (89)	339 (7)	16 (16)
Histology available, N (%)	54 (98)	70 (96)	221 (97)	319 (94)	13 (81)
Time to colposcopy (in days)	104	113	113	212	171
Number of cases (liberal)	12	2	11	3	1
Time to treatment (in days)	62	80	92	139	441
Number of cases (conservative)	11	1	8	0	0

[1] time interval between enrolment and first colposcopic evaluation
[2] time interval between first colposcopic evaluation and treatment

random assignment to colposcopy. At the outset we aimed to complete diagnostic assessments within 3 months of the enrollment visit. However, some cytology laboratories were experiencing delays of up to 6 months in providing cytological results. For this reason, the median time interval between screening tests and diagnostic procedures were 150 days overall (IQR: 80-274). This time interval did not vary by study arm (149 vs. 153; p value=0.73). However, in each study arm, this interval was shorter for women who had a positive screening test result: median of 106 days for participants with a positive screening test vs. 216 days for participants with negative screening test results (p value <0.001) in the Pap arm; and 111 vs. 212 days, (p value <0.001) in the HPV arm. Histology results were available for 95% of participants who underwent colposcopic evaluation, in each arm, and histological confirmation was as frequent for women with positive screening test results as those with negative screening test results (95% vs 96%). The median time interval between the first biopsy and the treatment was 70 days (Pap arm vs. HPV arm: 59 vs. 92 days; p value=0.23).

Through the procedure put in place to schedule participants for colposcopy, we were able to obtain colposcopic evaluation of 90% of participants with abnormal screening test results. However, as expected, participants with negative screening test results refused diagnostic evaluation more frequently (approximately half of those randomly flagged refused). Therefore, it became important to assess the potential differences between screen-negative participants who underwent diagnostic evaluation and those who did not, to ensure that those who did can be considered representative of those who did not.

Tables 11 and 12 provide an overview of the distribution of selected participants' characteristics by arm and centre. Among women testing negative on both screening tests, there was no meaningful difference on any of the sociodemographic characteristics between those who underwent colposcopy and those who did not. There was no statistically significant difference either except for participants randomized to the "Focus on HPV arm" in St. John's: participants who underwent colposcopy were slightly older than those who did not (median age of 46 vs. 43).

Table 12 provides the details on the cases that were identified. Among the 55 cases of CIN2 or worse that were identified on biopsy, there were 16 CIN2, 28 CIN3, 4 HSIL non specified, 3 adenocarcinoma in situ, 3 micro-invasive squamous carcinoma and 1 invasive squamous carcinoma. Among the 41 cases that were confirmed (conservative case definition) there were 9 CIN2, 25 CIN3, 1 HSIL non specified, 2 adenocarcinoma in situ, 3 micro-invasive squamous carcinoma and 1 invasive squamous carcinoma. Control histology (excision specimen or control biopsy) was available for 87% (47/55) of CIN2 or worse lesions identified on biopsy.

Table 11. Characteristics of Montreal screen negative participants, according to verification of disease

		Focus on Pap group			Focus on HPV group		
Verification of disease status by colposcopy:		Not required	Required but not completed	Completed	Not required	Required, but not completed	Completed
Characteristic[1]	Category	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Age	30-39	474 (30.9)	81 (35.4)	67 (32.7)	502 (32.2)	68 (32.1)	72 (33.0)
	40-49	553 (36.1)	76 (33.2)	67 (32.7)	549 (35.2)	82 (38.7)	75 (34.4)
	50-59	382 (24.9)	55 (24.0)	48 (23.4)	383 (24.6)	41 (19.3)	48 (22.0)
	60-69	123 (8.0)	17 (7.4)	23 (11.2)	124 (8.0)	21 (9.9)	23 (10.6)
Marital Status	Married/ in union	983 (65.1)	157 (70.1)	138 (68.7)	1 064 (69.1)	142 (68.3)	137 (64.3)
	Single	268 (17.7)	33 (14.7)	26 (12.9)	242 (15.7)	37 (17.8)	32 (15.0)
	Separated/ widowed	260 (17.2)	34 (15.2)	37 (18.4)	234 (15.2)	29 (13.9)	44 (20.7)
Contraception	None	640 (43.4)	98 (45.2)	83 (41.5)	674 (44.6)	91 (45.5)	94 (45.6)
	Hormonal	255 (17.3)	42 (19.4)	33 (16.5)	240 (15.9)	30 (15.0)	32 (15.5)
	Sterilization	288 (19.6)	40 (18.4)	46 (23.0)	310 (20.5)	37 (18.5)	42 (20.4)
	Barrier	219 (14.9)	28 (12.9)	29 (14.5)	214 (14.2)	32 (16.0)	27 (13.1)
	IUD	59 (4.0)	8 (3.7)	9 (4.5)	64 (4.2)	9 (4.5)	10 (4.9)
	Natural	12 (0.8)	1 (0.5)	0 (0)	9 (0.6)	1 (0.5)	1 (0.5)
Current smoker	Yes	342 (22.6)	56 (25.0)	51 (25.1)	345 (22.3)	41 (19.7)	42 (19.6)
	No	1 162 (76.6)	167 (74.6)	152 (74.9)	1 196 (77.3)	166 (79.8)	171 (79.9)
Self-report of ever having abn. Pap	Yes	333 (22.0)	48 (21.6)	47 (23.2)	311 (20.1)	36 (17.5)	49 (22.9)
	No	1 031 (68.2)	154 (69.4)	144 (70.9)	1 074 (69.6)	150 (72.8)	143 (66.8)
	Don't know	51 (3.4)	7 (3.2)	2 (1.0)	58 (3.8)	5 (2.4)	10 (4.7)

[1]Definition of variables and categories are the same as for Table 5. None of the differences in distribution reached statistical significance.

Table 12. Characteristics of St. John's screen negative participants, according to verification of disease status

		Focus on Pap group			Focus on HPV group		
Verification of disease status by colposcopy:		Not required	Required. but not completed	Completed	Not required	Required. but not completed	Completed
Characteristic	Category	N (%)	N %)	N (%)	N (%)	N (%)	N (%)
Age	30-39	988 (40.9)	85 (48.9)	34 (30.6)	1 025 (41.8)	70 (57.4)	35 (26.7)
	40-49	871 (36.1)	58 (33.3)	52 (46.8)	851 (34.7)	24 (19.7)	50 (38.2)
	50-59	425 (17.6)	28 (16.1)	22 (19.8)	471 (19.2)	25 (20.5)	33 (25.2)
	60-69	132 (5.5)	3 (1.7)	3 (2.7)	104 (4.2)	3 (2.5)	13 (9.9)
Marital Status	Married/ in union	1 967 (82.0)	138 (79.8)	89 (80.2)	2 005 (82.3)	101 (83.5)	112 (85.5)
	Single	159 (6.6)	17 (9.8)	10(9.0)	216 (9.0)	9 (7.4)	12 (9.2)
	Separated/ widowed	273 (11.4)	18 (10.4)	12 (10.8)	214 (8.8)	11 (9.1)	7 (5.3)
Contraception	None	1 207 (50.5)	82 (48.0)	52 (46.8)	1 171 (48.6)	60 (49.6)	73 (56.2)
	Hormonal	407 (17.0)	34 (19.9)	17 (15.3)	411 (17.1)	25 (20.7)	14 (10.8)
	Sterilization	554 (23.2)	33 (19.3)	36 (32.4)	583 (24.2)	25 (20.7)	37 (28.5)
	Barrier	186 (7.8)	18 (10.5)	5 (4.5)	203 (8.4)	11 (9.1)	6 (4.6)
	IUD	30 (1.3)	3 (1.8)	0 (0.0)	34 (1.4)	0 (0.0)	0 (0.0)
	Natural	7 (0.3)	1 (0.6)	1 (0.9)	8 (0.3)	0 (0.0)	0 (0.0)
Current smoker	Yes	400 (16.6)	30 (17.3)	16 (14.4)	380 (15.6)	21 (1.7)	14 (1.1)
	No	2 009 (83.3)	143 (82.7)	95 (85.6)	2 060 (84.4)	100 (82.6)	117 (89.3)
Self-report of ever having abn. Pap	Yes	667 27.7)	50 (28.9)	32 (28.8)	695 (28.5)	47 (38.8)	35 (26.7)
	No	1 524 (63.2)	112 (64.7)	70 (63.1)	1 514 (62.0)	65 (53.7)	83 (63.4)
	Don't know	116 (4.8)	5 (2.9)	8 (7.2)	120 (4.9)	7 (5.8)	6 (4.6)

[1]Definition of variables and categories are the same as for Table 5. Only the difference in age distribution reached statistical significance p=0.02 focus on Pap group, p<0.01 focus on HPV group.

Table 13. Selected characteristics of participants with biopsy proven CIN2 or worse

Case ID	Centre	Arm	Age	Pap result	HPV result	Lesion confirmed?[1]	Diagnosis
1	Mtl	Pap	45	AGC	Negative	No	CIN2
2	Mtl	Pap	40	HSIL	Positive	Yes	Inv.Sq.Ca.
3	Mtl	Pap	52	ASC-US	Negative	No[2]	CIN2
4	Mtl	HPV	38	Negative	Positive	Yes	CIN3
5	Mtl	HPV	30	Negative	Positive	Yes	CIN3
6	Mtl	Pap	31	ASC-H	Positive	Yes	CIN3
7	Mtl	Pap	42	HSIL	Positive	Yes	CIN3
8	Mtl	HPV	45	Negative	Positive	Yes	CIN2
9	Mtl	Pap	38	ASC-US	Positive	Yes	CIN3
10	Mtl	HPV	43	AGC	Positive	Yes	AIS
11	Mtl	HPV	33	Negative	Positive	Yes	CIN2
12	Mtl	Pap	35	Negative	Positive	Yes	CIN3
13	Mtl	HPV	66	Negative	Negative	No	CIN3
14	Mtl	Pap	37	HSIL	Positive	Yes	CIN3
15	Mtl	HPV	46	HSIL	Positive	Yes	Micro-inv.
16	Mtl	HPV	35	HSIL r/o inv.	Positive	Yes	CIN3
17	Mtl	Pap	44	Inv.Sq.Ca	Positive	Yes	Micro-inv.
18	Mtl	HPV	43	HSIL	Positive	Yes	CIN3
19	Mtl	Pap	39	Negative	Positive	Yes	CIN3
20	Mtl	Pap	31	LSIL r/o HSIL	Positive	Yes[3]	CIN3
21	Mtl	HPV	40	Negative	Positive	Yes[4]	CIN2
22	Mtl	HPV	34	ASC-US	Positive	Yes	CIN2
23	Mtl	HPV	30	ASC-US	Negative	Yes	CIN2/3
24	Mtl	HPV	48	AGC	Positive	Yes	Micro-inv.
25	Mtl	HPV	51	Negative	Positive	Yes	CIN3
26	Mtl	HPV	49	Negative	Positive	Yes	CIN3
27	Mtl	Pap	34	Negative	Positive	Yes[4]	CIN2
28	Mtl	Pap	44	ASC-US	Positive	Yes[4]	CIN3
29	Mtl	Pap	35	LSIL	Positive	Yes[5]	CIN3
30	Mtl	HPV	36	ASC-US	Positive	No	CIN 2
31	Mtl	Pap	31	ASC-US	Positive	No	CIN 2
32	Mtl	HPV	68	Negative	Negative	No[6]	CIN 2/3
33	Mtl	HPV	53	Negative	Negative	No[6]	CIN 2/3
34	Mtl	HPV	50	ASC-US	Negative	No[2]	CIN 2/3
35	Mtl	Pap	57	Negative	Negative	No[2]	CIN 2
36	Mtl	HPV	51	Negative	Positive	No[2]	CIN 3
37	SJ	Pap	34	Negative	Positive	Yes	CIN 3
38	SJ	Pap	32	Negative	Positive	Yes	CIN 2
39	SJ	HPV	42	AGC	Positive	Yes	CIN 3
40	SJ	HPV	46	Negative	Positive	No	CIN 3

Case ID	Centre	Arm	Age	Pap result	HPV result	Lesion confirmed?	Diagnosis
41	SJ	HPV	37	ASC-H	Positive	Yes	CIN 3
42	SJ	Pap	32	Negative	Positive	Yes	CIN 2
43	SJ	HPV	32	HSIL	Positive	Yes[5]	CIN 3
44	SJ	HPV	40	ASC-US	Positive	Yes	CIN 3
45	SJ	Pap	36	HSIL	Positive	Yes	CIN 3
46	SJ	HPV	55	Unsatisfactory	Positive	No	CIN 2
47	SJ	Pap	37	Negative	Positive	Yes	AIS
48	SJ	Pap	31	Negative	Positive	Yes[5]	CIN 3
49	SJ	Pap	33	Negative	Positive	Yes	CIN 2
50	SJ	HPV	33	Negative	Positive	No	AIS
51	SJ	Pap	35	Negative	Positive	No	CIN 2
52	SJ	HPV	34	Negative	Positive	Yes	CIN 2
53	Mtl	HPV	36	HSIL	Positive	Yes	CIN 3
54	Mtl	Pap	33	ASC-US	Positive	Yes[3]	CIN 3
55	Mtl	Pap	29	HSIL	Positive	Yes	CIN 3

Mtl: Montreal; SJ: St. Johm's
[1]Lesion confirmed on excision specimen or biopsy
[2]Participants had normal colposcopic exam, and ablative treatment without control biopsy
[3]Participants underwent LEEP on first colposcopy visit
[4] Participants had abnormal colposcopic exams, and ablative treatment without control biopsy
[5] Participants had biopsy before ablative treatment
[6] Participants had control biopsy and no treatment

4.4 Primary analysis

4.4.1 Comparison of Pap testing and HPV testing for screening cervical cancer and its high grade cervical precursors

Table 14 shows arm-specific sensitivity and specificity estimates. Using the conservative case definition, the verification-bias corrected sensitivity of Pap cytology (55.4%) was significantly lower than that of HPV testing (94.6%). The corrected specificity of Pap cytology (96.8%) was significantly, though only slightly, higher than that of HPV testing (94.1%), based on the conservative case definition. The sensitivity estimate of Pap, and the specificity estimates of both Pap and HPV based on the liberal case definition are very

Table 14. Arm-specific comparison of Pap and HPV testing to identify CIN2 or worse

Case definition[1]	Indices	Test[2]	Crude[3] estimates (95%CI)	Uncorrected[4] estimates (95%CI)	Corrected[5] estimates (95%CI)
Conservative	Sensitivity	Pap	57.1 (34.0-78.2)	57.1 (34.0-78.2)	55.4 (33.6-77.2)
		HPV	95.0 (75.1-99.9)	95.0 (79.6-100)	94.6 (84.2-100)
	Specificity	Pap	80.6 (77.4-83.6)	97.2 (96.7-97.6)	96.8 (96.3-97.3)
		HPV	60.9 (57.2-64.6)	94.1 (93.3-94.7)	94.1 (93.4-94.8)
	Positive predictive value	Pap	8.7 (4.6-14.7)	8.0 (4.0-13.0)	7.1 (4.8-10.3)
		HPV	6.6 (4.0-10.1)	6.0 (3.7-8.9)	6.4 (5.0-8.0)
	Negative predictive value	Pap	98.3 (96.8-99.2)	99.8 (99.7-99.9)	99.8 (99.7-99.9)
		HPV	99.8 (98.7-100)	100 (99.9-100)	100 (98.6-100)
Liberal	Sensitivity	Pap	57.7 (36.9-76.6)	57.7 (36.9-76.6)	43.4 (13.2-73.6)
		HPV	82.8 (64.2-94.2)	82.8 (64.2-94.2)	45.9 (18.9-72.9)
	Specificity	Pap	80.9 (77.7-83.9)	97.3 (96.8-97.7)	96.9 (96.4-97.4)
		HPV	61.1 (57.4-64.8)	94.2 (93.4-94.8)	94.2 (93.5-94.9)
	Positive predictive value	Pap	10.9 (6.2-17.3)	9.9 (5.8-15.4)	9.1 (4.7-16.7)
		HPV	8.3(5.4-12.1)	7.5 (4.9-11.0)	8.0 (5.6-11.3)
	Negative predictive value	Pap	97.9 (96.3-99.0)	99.8 (99.6-99.9)	99.6 (99.3-99.8)
		HPV	98.8 (97.3-99.6)	99.9 (99.8-100)	99.4 (99.1-99.5)

[1] Conservative: cases considered only if confirmed on the LEEP specimen or in the confirmatory biopsy when ablative treatment was used; liberal: all histologically confirmed CIN2, CIN3, AIS, or cervical cancers based on any of the ecto and endocervical biopsy specimens; See text for details

[2] Positivity defined as ≥ASC-US for Pap cytology and ≥1 pg HPV/ml for the HPV test.

[3] Crude estimates include only participants who underwent colposcopy (see section 3.8.2 for details).

[4] Uncorrected estimates assume all non verified are normal (see section 3.8.2 for details).

[5] Corrected for verification bias (see section 3.8.2 for details).

All estimates are in percentage.

similar to those obtained using the conservative definition. However, the sensitivity of HPV testing fell to 45.4% upon correction if the liberal less stringent case definition was used. The corrected and uncorrected estimates of sensitivity and specificity are very similar, regardless of the case definition used, except for the sensitivity estimate of HPV testing based on the liberal case definition (82.2% uncorrected estimate vs. 45.9% corrected estimates).

Using the liberal case definition, four cases were identified in participants testing negative on both screening tests compared to none via the conservative case definition. As expected, these four cases were very influential when extrapolating the expected numbers of cases in test-negative women, hence the important difference in corrected sensitivity between the two case definitions. Moreover, only one out of eight (12%) HPV-negative CIN2+ lesions on biopsy were also found in the excision specimen, compared to 68% (17/25) for Pap negative CIN2+ lesions (p value=0.01). Comparable proportions of HPV-positive (40/46, 87.0%) and Pap-positive (24/29, 82.8%) CIN2+ lesions were confirmed in the excision specimen (p value=0.7).

4.4.2 Impact of protocol violations and missing variables

Table 15 summarizes findings from the different analyses exploring the impact of protocol violations and missing variables. The "per protocol" strategy excluded from each arm participants who had the index test collected second and participants who refused to undergo colposcopic examination. Per protocol estimates are very close to the estimates in Table 14.

Table 15. Arm-specific comparison of sensitivity and specificity estimates by analyses strategies for protocol violations and missing variables

Case definition[1]	Indices	Test[2]	Main analysis[3]	Per protocol	Assumption for missing screening test	
					All positive	All negative
Conservative	Sensitivity	Pap	57.1 (34.0-78.2)	57.1	57.1	54.5
		HPV	95.0 (79.6-100)	94.7	95.0	94.3
	Specificity	Pap	97.2 (96.7-97.6)	97.3	95.7	97.2
		HPV	94.1 (93.3-94.7)	94.4	93.8	94.1
Liberal	Sensitivity	Pap	57.7 (36.9-76.6)	57.7	57.7	55.5
		HPV	82.8 (64.2-94.2)	84.6	82.8	82.4
	Specificity	Pap	97.3 (96.8-97.7)	97.4	95.8	97.3
		HPV	94.2 (93.4-94.8)	94.5	93.9	94.2

[1] Conservative: cases considered only if confirmed on the LEEP specimen or in the confirmatory biopsy when ablative treatment was used; liberal: all histologically confirmed CIN2, CIN3, AIS, or cervical cancers based on any of the ecto and endocervical biopsy specimens; See text for details
[2] Positivity defined as ≥ASC-US for Pap cytology and ≥1 pg HPV/ml for the HPV test.
[3] Uncorrected estimates (see section 3.8.2 for details).
All estimates are in percentage.

To investigate the impact of missing screening test results, two extreme situations were simulated. In the first situation, arm specific analyses were repeated assuming that all missing screening test results were positive and that all such participants were free of disease. In the second scenario, all participants with missing screening test results were assumed to have negative results and the prevalence of disease was assumed to be twice that of the study population. As can be appreciated in Table 15, these assumptions had minimal impact on estimates of sensitivity and specificity.

4.5 Exploratory Analysis

4.5.1 Sampling order interference

Table 16 shows the results for key indicators of test performance based on the order of cervical sampling. There was no statistically significant difference (p values 0.4-1) in performance of both tests, whether specimens were first or second in order of collection, in terms of proportion of positive tests, proportion of unsatisfactory smears, proportion of ASC-US smears, distribution of RLU categories, and sensitivity or specificity.

The confidence intervals around the differences between sensitivity and specificity estimates are wide. However, the confidence intervals around the differences between all other proportions are not statistically significant and exclude any difference greater than 2% (overall positivity, unsatisfactory smears, ASC-US smears, distribution of viral loads). Such tight confidence strengthen the assertion that sampling order does not influence test performance.

4.5.2 Impact of varying positivity thresholds and test combinations on screening performance

Since the order of sampling did not affect test performance we pooled the two study arms to investigate how different test combinations would perform (Table 17). As expected, increasing the positivity threshold for Pap and HPV DNA testing when used alone

Table 16. Analysis of sampling order interference

Screening test	Performance indicators [1]	Pap first	HPV first	Difference (95% CI)	P value
Pap cytology	Overall positivity	3.0	2.7	0.3 (-0.3; 0.9)	0.4
	Unsatisfactory smears	1.4	1.4	0.0 (-0.4; 0.4)	0.8
	ASC-US smears	1.9	1.8	0.1 (-0.4; 0.6)	0.8
	Crude sensitivity[2]	57.1	60.0	-2.9 (-33.0; 27.2)	1.0
	Crude specificity	80.6	82.7	-2.1 (-6.3; 2.1)	0.3
HPV test	Overall positivity	5.8	6.3	-0.5 (-1.4; 0.4)	0.4
	Distribution of viral load (RLU) levels (%)				0.8
	<0.75	91.8	91.1	0.7 (-0.4; 1.8)	
	0.75-0.99	2.0	2.3	- 0.3 (-0.9; 0.3)	
	1-1.99	1.1	1.3	-0.2 (-0.6; 0.2)	
	2-3.99	0.8	0.9	-0.1 (-0.5; 0.3)	
	4-9.99	0.8	0.7	0.1 (-0.2; 0.4)	
	10-39.99	1.2	1.2	0.0 (-0.4; 0.4)	
	≥40	2.0	2.1	-0.1 (-0.7; 0.5)	
	Crude sensitivity	100.0	95.0	5.0 (-4.6; 14.6)	0.5
	Crude specificity	61.1	60.9	0.2 (-0.5; 5.4)	0.9

[1] Positivity defined as ≥ASC-US for Pap and ≥1pg/ml for HPV test; screening indices based on the conservative case definition; RLU: Relative Light Unit (1 RLU~1pg HPV/ml)
[2] Crude estimates are based on screening test result and disease status for participants who completed colposcopy.

resulted in a decrease in sensitivity and colposcopy referrals. The triage of ASC-US Pap results by subsequent HPV testing resulted in somewhat reduced sensitivity compared to Pap alone at an ASC-US threshold, but yielded fewer referrals (1.6 vs. 2.9%). Triaging all

HPV positive women with a subsequent Pap test resulted in estimates similar to the inverse triage strategy. Interestingly this "reverse" triage strategy reached the highest positive predictive value (21.4%). Co-testing achieved 100% sensitivity while resulting in an 8% referral rate.

Table 17. Comparison of Pap cytology and HPV testing using combined study arms according to different positivity thresholds and test combinations[1]

Screening approaches	Positivity defined as	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)	Number of tests needed for screening	Referrals[2] (%)
Pap only	≥ASC	56.4	97.3	8.5	99.8	9,959	2.9
	≥LSIL	42.2	99.1	17.5	99.7	9,959	1.0
HPV only	≥1 pg/ml	97.4	94.3	7.0	100.0	9,959	6.1
	≥2 pg/ml	81.1	95.5	9.1	99.9	9,959	4.8
Pap screen followed by HPV triage	Triage of all ASC-US smears; HPV ≥ 1 pg/ml	53.8	98.7	14.9	99.8	10,145	1.6
HPV screen followed by Pap triage	Triage of all HPV ≥1pg/ml; Pap threshold ≥ASC-US	53.8	99.1	21.4	99.8	10,563	1.1
Pap and HPV co-testing	Pap ≥ ASC-US or HPV ≥1 pg/ml	100.0	92.5	5.5	100.0	19,918	7.9

[1]Verification bias corrected estimates based on the conservative case definition and using pooled data from 9959 women with available Pap and HPV results in both study arms; LSIL: low grade squamous intraepithelial lesion.[2]referrals for diagnostic evaluation

4.5.3 Impact of selected patients characteristics and laboratory on screening performance

The objective of the last exploratory analysis is to identify factors which may influence Pap and HPV testing sensitivity and specificity. The methodology is detailed in section 3.8.3.3. The variables in Tables 18 and 19 (for Pap and HPV, respectively) are potentially influential factors, and were considered by a backward selection process based on AIC. Tables 18 and 19 (included to illustrate the final models) summarize the coefficients and their associated p value for each of the Pap and HPV models.

Table 18: Modeling Pap sensitivity and specificity: variables, coefficients and statistical significance.[1]

Variable	Categories	Univariate analysis		Multivariate analysis	
		Coefficient	P value	Coefficient	P value
Age	Less than 50	*	*	*	*
	50 and over	-0.39	0.016	-0.46	0.006
Marital status	Married/living in union	*	*	*	*
	Single	0.55	0.001	0.34	0.051
	Separated, widowed	0.69	<0.0001	0.65	<0.0001
Contraception	None or non hormonal	*	*	*	*
	Hormonal	0.47	0.002	0.36	0.017
Pap laboratory	Laboratory group 1	*	*	*	*
	Laboratory group 2	-0.70	<0.0001	-0.73	<0.0001
	Laboratory group 3	-1.2	<0.0001	-1.19	<0.0001
Schooling	Primary school	*	*	*	*
	High school	-0.14	0.54	NA	NA
	College	-0.24	0.30	NA	NA
	University	0.04	0.82	NA	NA
Smoking	Current smoker	*	*	*	*
	Current non smoker	0.03	0.822	NA	NA

*referent category
NA not retained in the final model
[1]Estimates denote absolute increment (+) or decrement (-) in the parameter of interest relative to the performance at the level of referent.

Factors identified as statistically associated with sensitivity and specificity, and as such included in the final model, were similar for Pap and HPV: age, martial status, type of contraception, and laboratory performing the test. However, smoking was only retained in the HPV testing model. Age was retained in both models, but in the Pap model, the only important difference was between those less than 50 years of age, compared to those over 50. Marital status was also identified in both models. In the Pap model, coefficients were different for (1) married/in union, (2) single and (3) divorced. In the HPV model, the only difference was between (1) married/in union and (2) not married/in union. Contraception was identified in both models and the contrast was limited to hormonal contraception users vs. those not using hormonal contraception.

Table 19: Modeling HPV testing performance: variables, coefficients and statistical significance.

Variable	Categories	Univariate analysis		Multivariate analysis	
		Coefficient	P value	Coefficient	P value
Age	30-39			*	*
	40-49	-0.67		-0.61	<0.0001
	50-59	-0.92		-0.85	<0.0001
	60-69	-2.47		-1.31	<0.0001
Marital status	Married/living in union	*	*	*	*
	Not married nor in union	1.27	<0.0001	1.17	<0.0001
Contraception	None hormonal	*	*	*	*
	Hormonal	0.72	<0.0001	0.35	0.001
HPV laboratory	Laboratory 1			*	*
	Laboratory 2	-0.4	<0.0001	-0.28	0.003
Schooling	Primary school	*	*	*	*
	High school	0.03	0.86	NA	NA
	College	0.20	0.26	NA	NA
	University	0.34	0.04	NA	NA
Smoking	Non smoker	*	*	*	*
	Smoker	0.56	<0.0001	0.29	0.006

*referent category
NA not retained in the final model
[1]Estimates denote absolute increment (+) or decrement (-) in the parameter of interest relative to the performance at the level of referent.

Predicted Pap sensitivity was calculated for different subgroups (not shown in tables). The predicted values varied widely among subgroups, from a low of 32% in married women over 50 not using hormonal contraception in laboratory group 3, to as high as 87% in divorced/separated women under 50 using hormonal contraception in laboratory group 1. Across the same subgroups, predicted Pap specificity ranged from 94.1% to 97.8%. The range of the predicted HPV sensitivity estimates was smaller: from 90.2 (in non smoking, married women over 60 in laboratory group 2) to 99.6. However, the predicted specificity estimated of HPV testing showed a wider range than that of Pap: predicted specificity of HPV testing ranged from 69.8% to 98.6%.

Table 20 summarizes the predicted impact (in terms of absolute percentage points) of each variable on sensitivity and specificity estimates of Pap testing. Although contraception was selected because of statistical significance in the final multivariate model for Pap testing, its impact on sensitivity (7.9%) and specificity (-1.1%) of Pap testing is somewhat limited. The laboratory group has the largest impact on sensitivity and specificity, followed by the marital status and age variable. Sensitivity was on average 25.8% less in the laboratory group 3 compared to group 1; 14% more in divorced/separated participants compared to married participants; and 10% less in participants 50 and older, compared to those younger than 50.

The predicted impact of various factors on HPV testing performance is summarized in Table 21. From these estimates, we can appreciate that age and marital status have the largest impact on HPV testing sensitivity and specificity estimates. Although the impact

Table 20. Importance of spectrum effect of selected variables on Pap testing sensitivity and specificity

Variable	Referent	Contrast category	Univariate analysis		Multivariate analysis	
			Impact on sensitivity[1]	Impact on specificity	Impact on sensitivity	Impact on specificity
Age	Less than 50	50 and over	-9.6	0.9	-10.0	1.8
Marital status	Married[2]	Single	13.3	-1.5	7.7	-1.4
		Divorced[3]	16.2	-1.9	14.1	-2.5
Contraception	No hormonal[4]	Hormonal	11.1	-1.3	7.9	-1.1
Laboratory	Group1	Group 2	-14.8	2.7	-14.1	3.4
		Group3	-27.0	3.8	-25.8	4.6
Schooling	Primary school	High school	-3.5	0.3	NA	NA
		College	-5.8	0.5	NA	NA
		University	-10.4	0.1	NA	NA
Smoking	Non smoker[5]	Smoker	0.9	-0.1	NA	NA

Table 21. Importance of spectrum effect of selected variables on HPV testing sensitivity and specificity

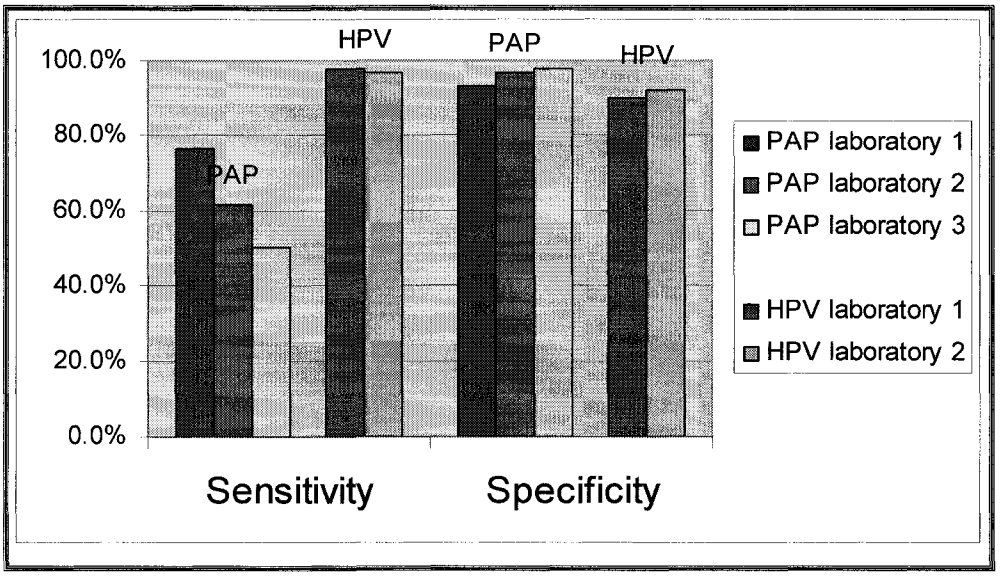
Variable	Referent	Contrast category	Univariate analysis		Multivariate analysis	
			Impact on sensitivity	Impact on specificity	Impact on sensitivity	Impact on specificity
Age	Less than 40	40-49	-1.7	3.7	-1.0	6.0
		50-59	-2.6	4.5	-1.6	7.8
		60-69	-4.5	5.5	-3.1	10.2
Marital status	Married	Not married	2.4	-7.6	2.8	-8.5
Contraception	No hormonal	Hormonal	1.4	-4.3	0.9	-2.7
Laboratory	Lab 1	Lab 2	-1.0	2.0	-0.7	2.1
Schooling	Primary school	High school	-0.1	0.7	NA	NA
		College	-0.5	0.9	NA	NA
		University	-0.7	1.7	NA	NA
Smoking	Non smoker	smoker	1.2	3.2	0.7	-2.2

[1]unweighted average difference in sensitivity between participants in the referent group compared to the contrast category. [2]inlcudes participants reporting living in as married. [3]Divorced and separated [4]includes users of non hormonal forms of contraception and those not using contraception. [5]Refers to current smoking status. Categories and definition are the same in table 21.

is limited to 3% in terms of sensitivity, it reached 10% for specificity. Contraception, laboratory and smoking were selected in the final multivariate model because of statistical significance, but their impact in terms of sensitivity and specificity is limited.

Evaluation of spectrum effect may inform clinicians, policy makers and resources planners on the expected performance of Pap and HPV testing in a specific population, defined by the variable categories included in the above described models. Although age and marital status are rather easily obtained in clinical practice, characteristics of laboratory performance are not. Indeed, in depth analysis which could inform on sensitivity and specificity parameters of any given laboratory cannot be routinely performed, as it would require diagnostic evaluation of screen-negative women. Given that this is the characteristic with the largest impact on Pap testing performance, it makes it impossible to accurately foresee the performance of Pap testing for a group of patients. On the other hand, as illustrated in Figure 10, HPV testing performance appears less variable across laboratories.

Figure 10. Variation of Pap and HPV testing performance by laboratory group



When trying to interpret the meaning of the marital status variable, which showed a statistical association and an important impact for both Pap and HPV testing performance, we had to address the issue of potential confounding (or intermediate) effect of sexual activity. It was hypothesised that marital status variable may have been a marker of sexual activity, more precisely the presence/absence of a new sexual partner in the last year. Unfortunately, such information was not available for the entire study population. This variable was available for Montreal participants only, and so it was decided to adopt the same LRM strategy to a restricted dataset including only Montreal participants, adding this new variable (new sexual partner in the last year) as a potential independent predictor.

Table 22 provides the coefficient of variables for univariate models and the final multivariate model of Pap testing sensitivity and specificity in Montreal, including a variable called "new sexual partner in the last year". In the final model, age, marital status and contraception were not retained. Only laboratory and "new sexual partner in the last year" were included in the final, best fitted model. Also, as mentioned in the methods section, the coefficient of the variable "presence or absence of disease" was compared in the initial "full" model and in the final model to ensure that the removal of variables had not created confounding.

The same strategy was used for HPV testing in Montreal. Table 23 provides the coefficients of the variables selected for the final model. Smoking and laboratory were not selected for the best fitted model when the variable "new sexual partner(s) in the last year" was added.

Table 22: Modeling Pap sensitivity and specificity in Montreal participants: variables, coefficients and statistical significance.

Variable	Categories	Univariate analysis		Multivariate analysis	
		Coefficient	P value	Coefficient	P value
Age	Less than 50	*	*	NA	NA
	50 and over	-0.33	0.086	NA	NA
Marital status	Married/in union	*	*	NA	NA
	Single	0.25	0.23	NA	NA
	Separated, widowed	0.29	0.18	NA	NA
Contraception	No hormonal	*	*	NA	NA
	Hormonal	0.37	0.07	NA	NA
Pap laboratory	Laboratory group 1			NA	NA
	Laboratory group 2	-0.71	<0.0001	-0.73	<0.0001
Schooling	Primary school	*	*	NA	NA
	High school	-0.08	0.80	NA	NA
	College	-0.22	0.46	NA	NA
	University	0.04	0.89	NA	NA
Smoking	Current smoker	*	*	NA	NA
	Current non smoker	0.31	0.15	NA	NA
New partner	None	*	*	*	*
	One or more	0.78	<0.0001	.80	<0.0001

*referent category. NA not retained in the final model

Table 23: Modeling HPV sensitivity and specificity in Montreal participants: variables, coefficients and statistical significance.

Variable	Categories	Univariate analysis		Multivariate analysis	
		Coefficient	P value	Coefficient	P value
Age	30-39	*	*	*	*
	40-49	-0.89	<0.0001	-0.74	<0.0001
	50-59	-0.99	<0.0001	-0.64	0.001
	60-69	-1.39	<0.0001	-1.02	<0.003
Marital status	Married/in union	*	*	*	*
	Not married/in union	1.07	<0.0001	0.76	<0.0001
Contraception	No hormonal	*	*	*	*
	Hormonal	0.79	<0.0001	0.43	0.005
Schooling	Primary school	*	*	NA	NA
	High school	-0.22	0.40	NA	NA
	College	0.22	0.35	NA	NA
	University	0.39	0.08	NA	NA
Smoking	Non smoker	*	*	NA	NA
	Smoker	0.34	0.015	NA	NA
New partner	None	*	*	*	
	One or more	1.51	<0.0001	1.01	<0.0001

*referent category. NA not retained in the final model

Table 24 summarizes the predicted impact of each variable on sensitivity and specificity for Pap and HPV testing. Only laboratory group and having a new sexual partner in the last year influenced sensitivity and specificity of Pap testing in the Montreal population (compared to age, marital status and contraception which were selected in the final model in the complete study population when sexual partners in the last year was not considered). The magnitude of the impact of those 2 variables is similar: 13.5-14.8% (in absolute terms) on sensitivity, and 3.6-3.9% on specificity.

In the Montreal population, age, marital status, contraception, and new partner were included in the final multivariate model for HPV testing. Age and having a new sexual partner in the last year were identified as having the largest impact on HPV testing performance, particularly on specificity, up to 8.8% (Table 25).

Decreased specificity is the main reason why HPV testing has not been incorporated in primary screening activities. Various triage strategies, which could increase specificity, are under investigation. However, given the differential impact of certain variables on Pap and HPV testing sensitivity and specificity, we hypothesised that certain subgroups of women may already derive benefits from HPV testing, namely that in certain subgroups HPV testing would have a higher sensitivity than Pap while having the same specificity.

To test this hypothesis, logistic regression modelling was done, including only 2 variables, categorized in the same manner for the Pap and HPV models. The

Table 24. Importance of spectrum effect of selected variables on Pap testing sensitivity and specificity, Montreal study population

Variable	Referent	Contrast category	Univariate analysis		Multivariate analysis	
			Impact on sensitivity[1]	Impact on specificity	Impact on sensitivity	Impact on specificity
Age	Less than 50	50 and over	-7.8	1.0	NA	NA
Marital status	Married[2]	Single	5.7	-0.9	NA	NA
		Divorced[3]	6.5	-1.0	NA	NA
Contraception	Non hormonal[4]	Hormonal	8.0	-1.3	NA	NA
Laboratory	Group1	Group 2	-14.2	2.7	-13.5	3.6
Schooling	Primary school	High school	-1.7	0.3	NA	NA
		College	-5.0	0.7	NA	NA
		University	0.8	-0.1	NA	NA
Smoking[5]	Non smoker	Smoker	6.9	-0.9	NA	NA
New partner[6]	None	I or more	16.8	-3.1	14.8	-3.9

Table 25. Importance of spectrum effect of selected variables on HPV testing sensitivity and specificity, Montreal study population

Variable	Referent	Contrast category	Univariate analysis		Multivariate analysis	
			Impact on sensitivity	Impact on specificity	Impact on sensitivity	Impact on specificity
Age	Less than 40	40-49	-3.0	6.0	-1.7	7.0
		50-59	-3.5	6.4	-1.4	7.2
		60-69	-6.1	7.8	-2.7	8.8
Marital status	Married	Not married	3.0	-7.0	2.2	-5.6
Contraception	No hormonal	Hormonal	2.1	-5.8	1.3	-3.9
Smoking	Non smoker	Smoker	1.1	-2.2	NA	NA
Schooling	Primary school	High school	-0.8	1.0	NA	NA
		College	0.7	-1.3	NA	NA
		University	1.1	-2.4	NA	NA
New partner	None	I or more	3.6	-12.7	3.0	-8.0

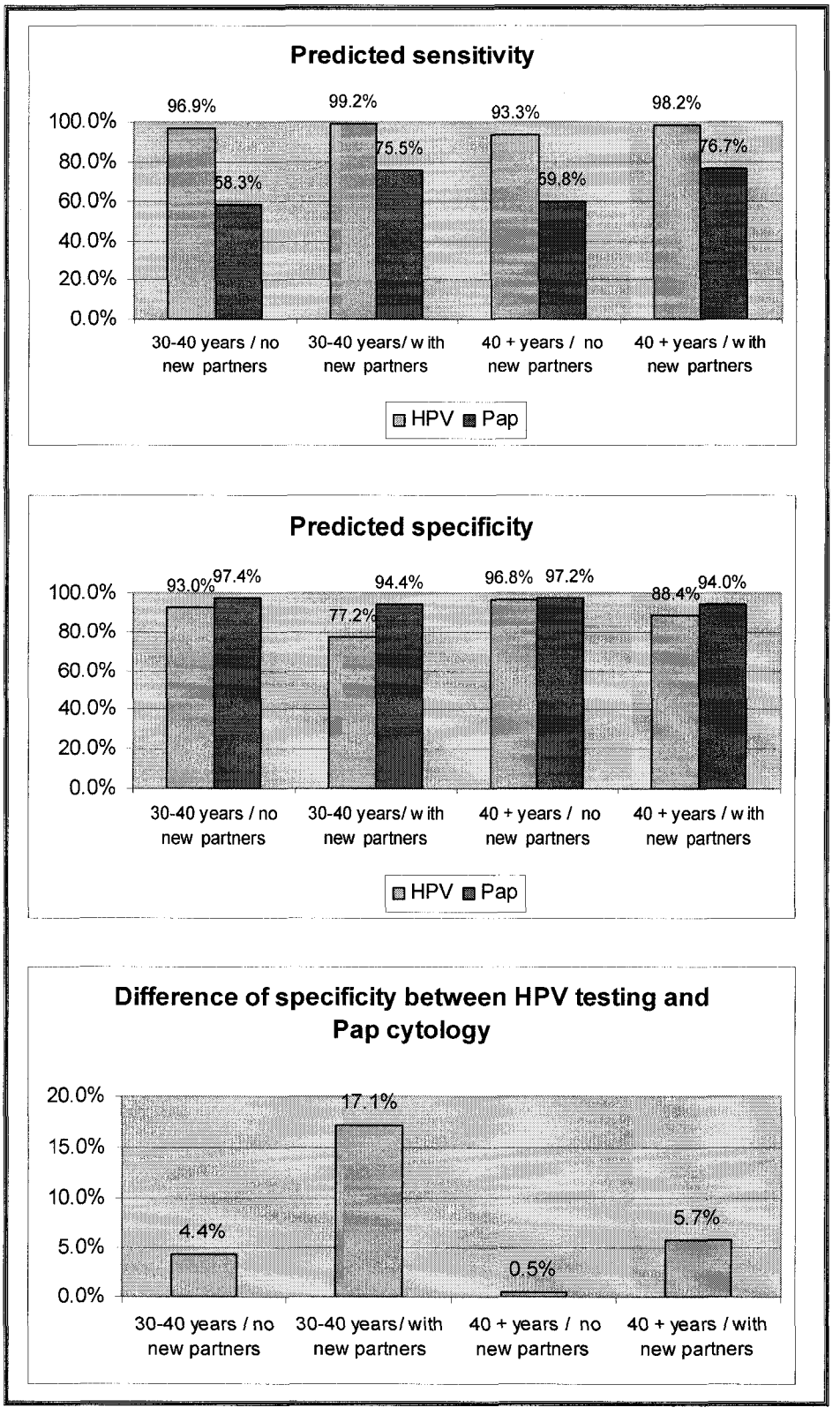
[1]unweighted average difference in sensitivity between participants in the referent group compared to the contrast category. [2]inlcudes participants reporting living in as married. [3]Divorced and separated [4]includes users of non hormonal forms of contraception and those not using contraception. [5]Refers to current smoking status. [6]refers to presence of a new sexual partner in the last year. Categories and definition are the same in table 23.

variables selected were age and new sexual partner. New sexual partner was selected because of its impact on both Pap and HPV testing performance. Age was selected because of its important impact on HPV testing specificity. Although laboratory is an important predictor for Pap performance, the performance of a given cytology laboratory is circumstantial to the study's setting, and thus it cannot be used to select women who may benefit more from one test or another in routine screening. Marital status and contraception were not selected, because their impact on sensitivity and specificity were limited

Thus, in summary, the third set of models included participants from Montreal, and only age and "new sexual partner in the last year" were considered. In this last set of models, the impact of age was only visible in women under 40 years of age compared to those 40 and over. In this analysis, predicted Pap sensitivity varied from 58% to 78%, compared to HPV sensitivity which varied from 97% to 99%. Specificity ranged from 94% to 97% for Pap, and from 77% to 97% for HPV testing (Figure 11).

Predicted HPV testing was at least 20% more sensitive than Pap testing in all subgroups. However, as illustrated in Figure 11, the differences in specificity varied widely among subgroups. Interestingly, in women over 40 reporting no new sexual partner in the last yet, the predicted difference in specificity was only 0.5%. Yet, the predicted gain in sensitivity in the same subgroup was 33.5%.

Figure 11. Predicted sensitivity and specificity estimates of Pap and HPV testing, by age and "new partner"



5 DISCUSSION

5.1 Summary and significance of findings

This thesis summarises the results of the first RCT designed to compare Pap cytology with HPV testing as standalone tests for screening high grade cervical cancer precursors. The ability to recruit over 10,000 participants in community practices in just over 2 years can certainly be attributed, at least in part, to the enthusiasm physicians and women felt toward improving cervical cancer screening.

Overall, 2.9% of participants had a positive Pap test, and 6.0% had a positive HPV test. The proportion of positive tests (Pap and HPV) decreased with increasing age. The proportion of positive HPV tests increased with the severity of the Pap test results. Almost all participants (14/15) with the most abnormal smears (≥HSIL) were HPV positive. The numbers of prevalent CIN2 or worse lesions was low: 54 by the liberal case definition and 41 cases by the conservative case definition. These findings translate into prevalence estimates of 5.3 and 4.3 per 1000 women, respectively (sections 4.2 and 4.3).

5.1.1 Summary and significance of main findings

We have reported estimates based on 2 case definitions: liberal and conservative. However, we believe that the conservative case definition is the most accurate and most relevant to clinical practice, as detailed in section 5.2.4.

Our findings based on the conservative case definition concur with those of other studies [Cuzick 1999a, Kuhn 2000, Ratnam 2000, Schiffman 2000, Schneider 2000, Belinson 2001, Blumenthal 2001, Clavel 2001, Kulasingam 2002, Belinson 2003, Cuzick 2003, Petry 2003, Salmeron 2003, Agorastos 2005, Sankaranarayanan 2004, Bigras 2005, Ronco 2006] showing that HPV testing is more sensitive than Pap cytology (39.2% difference, conservative case definition) for screening cervical cancer and its high grade precursor lesions. The specificity of HPV testing was only 2.7% lower than that of Pap testing. The PPV of HPV testing was slightly lower than that of Pap (0.7% difference) with confidence intervals largely overlapping. Not surprisingly, owing to the low prevalence of lesions in the two population samples, the NPV of both tests was above 99% (Table 14).

Although there are no data available from randomized trials designed to compare Pap and HPV as standalone tests, other observational studies have provided information, either through other designs (split sample studies) with the primary aim of comparing Pap to HPV or in secondary analysis of studies designed to address primarily other questions. This information is summarized in Table 26. All studies except the one by Coste [2005], found HPV testing (threshold of 1pg/ml) to be more sensitive than Pap testing (ASC-US threshold) to identify CIN2 or worse lesions. The fact that similar results were found in a randomized design adds confidence to the results of previous observational studies. It should be noted that the 3 studies which found HPV testing to be 80% sensitive or less [Blumenthal 2001, Coste 2003, Sankaranarayanan 2004] had colposcopy performed on all participants, and histological diagnosis was not required for definitive diagnosis. As such, it is possible that the apparent lower sensitivity may in fact be due to errors in diagnosis [Arbyn 2006].

Table 26. Pap and HPV testing for primary screening of cervical cancer precursor: design characteristics and results of published trials

Study	Country	Sample Size	Blinding[1]	Histology obtained	Verification bias correction	Sensitivity (%)		Specificity (%)		Referrals (%)	
						HPV	Pap	HPV	Pap	HPV	Pap
Belinson, 2001	China	1997	NS	90%	By design	95	94	85	78	18	25
Cuzick, 1999a	U.K.	2988	NS	NS	Not done	95	86	NA	90	7	6
Kuhn, 2000	South Africa	2944	Histology	NS	Not done	88	78	82	97	22	8
Ratnam, 2000 [2]	Canada	595	Histology	NS	By analysis	82	40	94	97	5	4
Schiffman, 2000	Costa Rica	8554	Histology	NS	Not done	88	78	89	94	12	7
Blumenthal, 2001	Zimbabwe	2073	No	NS	By design	80	44	61	91	43	13
Clavel, 2001[4]	France	5671	NS	NS	Not done	100	75	89	95	12	6
Belinson, 2003	China	8497	Histology	NS	Not done	97	88	78	81	24	22
Cuzick, 2003	UK	10358	No	50%	Not done	97	77	93	96	8	5
Kulasingam, 2002	USA	4075	Histology	NS	By analysis	91	61	73	82	29	19
Petry 2003	Germany	7908	Histology	54%	By analysis	97	46	95	98	5	2
Schneider, 2000	Germany	4761	NS	74%	By analysis	80	20	91	99	8	1
Salmeron, 2003	Mexico	7732	No	42%	Not done	93	59	91	98	9	2
Agostaros, 2004	Greece	1296	Histology	NS	Not done	75	50	97	98	3	2
Sankaranarayanan, 2004	India	18085	Yes	20%	By design	68	66	93	95	7	6
Coste, 2005	France	1757	NS	NS	By design	64	72	86	94	3[2]	12
Bigras, 2005	Switzerland	13865	NS	80%	By analysis	97	92	59	97	8	4
Ronco, 2006	Italy	16706[3]	No	NS	Not done	97	74	93	82	7	5

NS: Not Specified; Sensitivity for high grade lesion at ASC-US (or LSIL is ASCUS unavailable) threshold for Pap and 1 pg/ml for HPV;
[1]Blinding: histology=only pathologists blinded; colposcopists were not blinded. Yes: Both colposcopists and pathologists were blinded.
[2] % referrals estimate for HPV testing from de Cremoux 2003;
[3] Sample size of entire study is larger, but comparison of HPV vs. Pap done on sample size of 16706

In their meta-analysis focusing on the comparison of Pap with HPV testing for primary screening of cervical cancer and its precursors, Arbyn et al. concluded that HPV testing is 23% more sensitive than Pap testing, and 6% less specific [Arbyn 2006]. A review limited to observational studies conducted in North America and Europe also reached the conclusion that HPV testing is superior to Pap testing, in terms of sensitivity (96% vs. 53%; uncorrected estimates), with an associated loss of 6% in specificity [Cuzick 2006a].

The consistency of the finding that HPV testing is more sensitive than Pap testing to identify CIN2 or worse lesions supports a change in paradigm for cervical cancer screening from cellular to virological tests. The following section discusses various options to meet the biggest challenge associated with this change: to benefit from the high sensitivity of virological tests, while keeping referrals as low as possible in order to avoid the morbidity and costs of unnecessary procedures.

5.1.2 Summary and significance of exploratory analysis findings

We found no evidence that Pap or HPV performance was influenced by the order of specimen collection when multiple cervical samples are obtained, e.g., in the case of Pap and HPV co-testing (Table 16). The reading of a Pap smear may be made more difficult in the presence of blood, inflammatory cells or mucus [Davey 1996, Hidelsheim, 1998, Solomon 2002]. Sampling the cervix for another test before collecting the Pap test could, in theory, reduce the number of undesired elements by removing excess mucus, or increase the number of red blood cells due to the micro- trauma of the first sampling. If

either happened, it did not reflect on the performance indicators that are most likely to be affected by specimen quality and quantity.

There is no literature available on the effect of collecting other cervical samples before collecting one for HPV testing by HC2. All published studies comparing Pap to HPV testing either (1) collected a Pap sample first and a second one for HPV testing, or (2) collected one specimen, prepared a Pap smear first and used residual material for HPV testing. Consequently, there is no data on the performance of HPV testing in conditions similar to those that would be found if HPV testing were used as a standalone screening test, ie the specimen collected first, and used entirely for processing. PCR, a common assay used for HPV DNA testing in research settings, may be inhibited in the presence of blood caused by the micro-trauma of the first sampling. Also, taking another sample prior to the HPV one could, in theory, reduce the amount of virions and as such cause false negative results. It was thus important to assess if HPV testing performed differently when collected first. We found no effect in terms of overall positivity or in distribution across viral load categories.

The above findings enabled us to pool data from both study arms to explore different testing scenarios (Table 17). Since viral load has been associated with severity of disease [Sun 2001, Sherman 2003, Snijders 2003, Snijders 2006], raising the threshold for HPV positivity has been suggested as a mean to increase specificity and decrease referrals. We found that raising HPV positivity threshold from 1 to 2 pg/ml would reduce referrals for colposcopy from 6.1% to 4.8% while keeping sensitivity (81.1%) greater than that of Pap cytology at an ASC-US threshold (sensitivity 56.4% both arms, referrals 2.9%). However,

this strategy still has a relatively low PPV at 9.1%. Others reported minimal impact on sensitivity and specificity when raising the HPV positivity threshold from 1 pg/ml to 2 or 3 pg/ml [Cuzick 1999a, Clavel 2001, Belinson 2003, Cuzick 2003, Ronco 2006].

Our results support previous findings suggesting that HPV triage of ASC-US smears is nearly as sensitive as immediate colposcopy, while keeping referrals low [Cox 1995, Manos 1999, Solomon 2001, Zielinski 2001, Kim 2002, ALTS 2003, Arbyn 2004, Bais 2005, Berkhoff 2006, Ronco 2007a]. Indeed sensitivity estimates of Pap only vs. Pap with HPV triage of ASC-US smears were very similar (56.4% vs 53.8%) but the triage strategy was associated with fewer referrals (Table 17).

An alternative algorithm using a more sensitive test (HPV) as a primary screening tool followed by triage of those who are HPV positive with a more specific test (Pap) has gained increased support in the literature [Cuzick 2006]. Our post-hoc assessment of this strategy indicated a much lower sensitivity than HPV testing used alone (53.8% vs. 97.4%). However, cytotechnicians were blinded to HPV test results in CCCaST. In a true reflex triage situation cytotechnicians would be made aware that the slides to be read are from HPV positive women, which would likely lead to more meticulous assessment of smears within a much reduced case load [Franco 2006]. Although compelling, this hypothesis remains to be tested. A large RCT (200,000 participants) is underway in Finland comparing primary HPV screening with positive tests triaged by Pap cytology, to conventional pap testing. Results should be available in 2009 [Davies 2005, Kotaniemi-Talonen 2005]. It is interesting to note that in CCCaST this strategy had the highest positive predictive value, at 21.4%.

Co-testing, an acceptable option in cervical cancer screening in the United States [Wright 2004, Smith 2006], reached 100% sensitivity in CCCaST. In fact, sensitivities of 100% have been reported in all trials addressing the performance of co-testing [Ratnam 2000, Clavel 2001, Cuzick 2003, Petry 2003, Ronco 2006]. The robustness of this finding is strengthened by the fact that no single high grade lesion has been reported in the 3162 colposcopies reported in various trials in Pap-negative HPV-negative participants [CCCaST, Schiffman 2000, Belinson 2001, Kulasingam 2002, Cuzick 2003, Petry 2003, Bigras 2005]. Only Blumenthal [2001] reported some high grade CIN in Pap negative HPV negative women. However, colposcopy was performed on all participants biopsies were not required. Over 40% of women were HPV and some were immunosuppressed. Specificities of population characteristics and diagnostic ascertainment render the findings less than conclusive.

The cost-effectiveness of co-testing will need further evaluation, since it only marginally improves sensitivity compared to HPV testing alone, while doubling the number of screening tests and substantially increasing colposcopy referrals.

Although sensitivity and specificity are largely unaffected by disease prevalence, these performance indicators may vary with patient characteristics [Ransohoff 1978, Armitage 1994, Mulherin 2002, Zhou 2002]. The term spectrum effect is used to describe the heterogeneity of test performance indicators across patient subgroups with different characteristics [Ransohoff 1978, Mulherin 2002]. This situation may be particularly problematic when new screening tests are applied to a high risk population, already

selected for diagnostic evaluation, such as when HPV testing was studied in colposcopy clinic populations [Coste 2003]. Performance evaluation obtained in such studies is often not generalizable to the general population for which the screening test is intended. However, even when the study design correctly includes the population which is the ultimate target for the introduction of the new screening test, it is important to assess how test performance may vary across population subgroups.

Presenting relevant participant characteristics, restriction and stratified analysis are all strategies that have been proposed to deal with spectrum effect. However, with a large number of candidate spectrum effect variables, logistic regression modeling is more efficient. Such analysis takes advantage of the entire data available and makes it possible to investigate the effect of several covariate including continuous ones [Coughlin 1992, Mulherin 2002].

Our aim with spectrum effect analysis was two-fold. Firstly, we wanted to identify which factors influenced Pap and HPV test performance, and quantify this effect. Secondly, we wanted to investigate if certain subgroups of women may derive obvious benefits from switching from Pap to HPV testing for cervical cancer screening, that is identify subgroups of participants for which there would be an important gain in sensitivity but a limited loss in specificity.

Pap testing was introduced in clinical practice before formal evaluation of the effectiveness of screening methods became standard of evidence. Ecological and case-control studies have since provided strong evidence of its impact on cervical cancer

incidence and mortality [Franco 2002]. Although there is published data on women and health care worker characteristics which influence Pap uptake, and on laboratory organization characteristics which influence Pap performance [Peters 1988, Brinton 1994, Jepson 2001, Breen 2005, Forbes 2007, Tacken 2007], there is limited data on patient characteristics which may impact on test performance. In their meta-analysis, Nanda et al., have underlined the wide range of reported Pap sensitivities in the literature, and have attributed observed differences to deficient study design, heterogeneity in laboratory quality assurance measures, and differences in disease prevalence [Nanda 2000]. The latter essentially refers to the fact that most studies on Pap performance were conducted in colposcopy clinic in so called "high-risk" patients. No study was identified specifically addressing which patient characteristics may influence Pap testing. Studies on the use of HPV testing in cervical cancer screening are relatively recent. Early on, age was identified as having a negative impact on specificity [Cuzick 1995, Cuzick 1999a, Schifman 2000, Wright 2004]. For this reason, further studies have limited participation to women over 30. There is no study looking at other patient characteristics which may influence HPV testing performance.

We explored the effect of different patient characteristics on Pap and HPV performance. We assessed the same characteristics for Pap and HPV models in order to compare any effect that was identified. The variables included for consideration were either (1) known to influence test performance, such as age and laboratory conducting the analysis, (2) or risk factors for HPV infection and cervical cancer, such as smoking, markers of sexual activity (marital status, new sexual partner in the last year), type of contraception and

marker of socioeconomic status (education) [Castellsague 2003, Kamangar 2006, Trottier 2006].

In our analysis, schooling and smoking were not associated with Pap or HPV testing performance. Smoking reached statistical significance for HPV testing only if presence of a new partner was not included in the model, and then only had a limited impact. We found that laboratory and markers of sexual activity (marital status or new sexual partner in the last year) had the largest impact on Pap performance, mainly on sensitivity. Age and markers of sexual activity had the largest impact on HPV testing performance, but mainly on specificity (Tables 20 and 21).

The fact that the laboratory performing the test has such a large influence on Pap testing performance is worrisome, given that quality assurance measures have been implemented across Canada for a long time and would have been expected to reduce the heterogeneity of performance. Moreover, it is very difficult to know the performance of any given laboratory in routine practice and use this knowledge for clinical decision making. On the other hand, the effect of laboratory on HPV testing performance appears very limited (1% on sensitivity and 2% on specificity, Table 21). This can probably be largely explained by the fact that HC2 is a rather simple, highly standardized and partially automated procedure, compared to cytology which is by nature subjective. Also, the sampling errors are an important limitation of Pap testing [Stuart 1997, Sadler 2004a, Leyden 2005, Spence 2007]. If the correct part of the cervix (the transformation zone) in not sampled, diseased cells may not be available for diagnosis. However, HPV genital infection is most often a "regional" infection and sampling of a precise area of the cervix

may not be as critical as for Pap testing. This is the reason why self sampling can be considered a viable option for HPV testing [Bais 2007, Petitgnat 2007]. Nevertheless, these interesting findings will need to be replicated in other studies including more HPV laboratories, since only 2 HPV laboratories participated in CCCaST.

The effect of age, marital status and contraception on Pap testing performance disappeared when we controlled for the presence of a new sexual partner in the last year (Table 22). However, it is difficult to reach firm conclusions, since only Montreal participants were included in the model with the variable "new sexual partner", as opposed to the model without that variable which included all CCCaST study participants. On the other hand, age and new sexual partner were both independently associated with HPV testing performance (Table 23). There is a plausible explanation for why having a new sexual partner in the last year affected the performance of Pap and HPV testing. Indeed, having a new sexual partner is the most important risk factor for acquiring a new HPV infection [Bauer 1993, Moscicki 2001, Castle 2005, Dunne 2007]. The HPV DNA of this new infection will be detected by HPV testing, and the cytological changes associated with it will be detected by Pap testing. However, only persistent infection increases the risk of significant cervical disease [Koutsky 1992, Ho 1995, Remmink 1995, Chua 1996, Wallin 1999, Schlecht 2001, Kjaer 2002]. As such, having a new partner in the last year will not have a significant impact on high grade cervical precancerous risk. By having a different impact on risk of having a positive test and risk of having disease, this variable affects test performance.

The mechanism by which age may affect HPV testing performance independently from markers of sexual activity is less straightforward. Of course, age is a marker for "time" and older women are more likely to have HPV infections which have persisted for longer periods, compared to younger women. Thus, in the presence of a positive HPV test, the infection is more likely to be a persistent one, and more likely to have been the source of the underlying lesion. Age is also related to ectopy (presence of a glandular component on the ecto-cervix) which is itself associated with HPV type distribution in the cervix and adjacent areas of the upper vagina [Castle 2006]. Consequently, age-related effects on the anatomy of the cervix may have differentially affected detection of HPV types that are identified in the HC2 assay.

When predicting Pap and HPV sensitivity and specificity in subgroups of women defined by age (younger or older than 40) and sexual partner in the last year (presence or absence of a new sexual partner in the last year), we gained interesting insights in potential use of HPV testing. Older women reporting no new sexual partner in the last year would benefit from an important increase of sensitivity associated with HPV testing (33%), while still benefiting from a specificity of 97% (Figure 11). It is the only subgroup for which the PPV of HPV testing was not lower than that of Pap testing (HPV 7% vs. Pap 5%). Although this information may be of limited benefit in the context of organized programs sending invitation letters on the basis of information easily accessed through administrative databases (age typically), it may be more useful in opportunistic settings, where each health care provider is left to decide when and what test to perform for each woman. Of course, guidelines for screening should be simple and universal even in the case of opportunistic screening. However, the heterogeneity of performance of screening

tests across subgroup of patients sometimes makes risk-stratified algorithms essential, as is the case for breast and colorectal cancer screening [Smith 2006].

5.2 Methodological considerations

Although there has been mounting evidence since the late 1990s that HPV testing is more sensitive than Pap testing to screen for high cervical cancer precursors, HPV testing has not been implemented in routine population screening, except in the United States. Although concerns about costs and questions regarding patient management certainly contributed to this restraint in implementing HPV based screening, two other factors were of utmost importance: the lower specificity of HPV testing and methodological limitations of studies suggesting that HPV testing was more sensitive. Indeed, reviews and technology assessment reports concluded that methodological considerations precluded firm conclusions on the value of HPV testing and that studies with more sound design characteristics were needed, ideally RCTs, before policy changes could be recommended [Cuzick 1999b, Noorani 2003, Arbyn 2006, Hulstaert 2006, PAB 2007]. Various triage and/or restriction of target population which could address the decrease in specificity associated with HPV testing have been discussed in the previous section. The following section focuses on CCCaST design specificities which attempted to address design limitations (detailed in section 1.3) of studies on HPV testing for primary cervical cancer screening.

5.2.1 Design

Randomizing participants to Pap or HPV testing would have been the most straightforward design to compare test performance. However, the paucity of data on HPV testing made it ethically impossible to withhold Pap testing from study participants at the onset of CCCaST. In consequence, we decided to include both tests in each arm, but randomize the order of sampling. This design enabled us to report, for the first time, how HPV testing would perform if used as a standalone test. Since all previously published studies had collected a Pap test first and no one had investigated the effect of consecutive collection, the actual performance of HPV testing as a standalone test remained in doubt.

Our main objective was to compare sensitivity, specificity and predictive values of Pap and HPV testing. To accomplish this we decided to investigate with the gold standard (colposcopy and biopsy) all participants with positive tests and a random sample of participants with negative tests. Since HPV testing can be expected to approximately double the number of colposcopy referrals compared with Pap testing, investigating women only on the basis of a positive index test result would have created in imbalance in colposcopy referrals, which can induce bias. Indeed, whatever the intrinsic value of a screening test, sending more women for disease verification will increase the likelihood of finding disease and bias sensitivity estimates upward [Franco 1999b]. Our design avoided this pitfall as we had similar proportions of disease verification between our 2 study arms (Figure 8).

However, when high grade cervical cancer precursor lesions were identified, they had to be treated according to prevailing clinical guidelines, without consideration of the screening test which led to its identification. For this reason, even if we followed our participants for many years, this study design will not inform us on the differential impact of Pap and HPV testing on cervical cancer incidence or mortality [Davies 2005, Ronco 2006]. Other larger trials, some with passive follow-up within organized programs and record linkage to tumor registry data and mortality databases, have since been designed to address this issue [Kotaniemi 2005, Ronco 2006, Coldman 2007].

.

5.2.2 Population

Our recruitment strategy was designed to include participants who were representative of women undergoing routine cervical cancer screening in Canada. The lack of a formal organized screening program was no deterrent to successful accrual. As others, we found that working with community medical practices provided an efficient strategy to make the study available to the population who would ultimately be targeted for the intervention [Sellors 2002]. Less than 10% of the women approached declined to participate. Recruitment took place in university-affiliated clinics and private practices, in family medicine and gynaecology practices of different sizes that focus on primary care. This strategy enabled us to recruit the first large North American population of women, who had had previous access to cervical cancer screening.

It is impossible to know if our findings would apply to the general population if an organized program with an increased coverage were put in place. Women who do not

avail themselves of cervical cancer screening are often different from those who do [Brinton 1994, Miller 1994, Breen 2005]. However, it is reassuring to note that HPV testing performance has recently been investigated within more organized programs and has shown similar performance [Ronco 2006, Table 26].

5.2.3 Screening tests

Both Pap and HPV tests were collected and processed in the usual community setting where screening takes place. Although we performed a quality control of HC2, we did not use this information to change physician collection or laboratory handling of the samples. Not only were technicians not aware of other test results, but they were not even aware of participation in a study. This ensures that our results truly represent what can be expected of these tests in everyday practice. This is different from trials in communities where screening is not usually available and which, by design, have tests collected and sent elsewhere for expert reading. The evaluation of screening tests in such circumstances is closer to the ideal rather than to the usual performance.

5.2.4 Diagnostic evaluation and case definition

The availability of a gold standard is no doubt essential to determine the validity of diagnostic or screening tests [Zhou 2002]. For example, failure to apply the gold standard in many studies on cervical cancer screening was the main reason why there was a paucity of data on which to base systematic reviews [Cuzick 1999b, Noorani 2003, Hulstaert 2006, Arbyn 2006, PAB 2007]. In the case of cervical cancer precursor

diagnosis, disease verification is a 2-step process: visual examination (colposcopy) and histological assessment of biopsy specimens. We designed a colposcopy protocol to standardize disease verification as much as possible, and thus reduce misclassification and bias. As colposcopic impression alone has been shown to be an unreliable diagnostic tool for high grade cervical cancer precursors [Mitchell 1998a, Olaniyan 2002, Guido 2003, Benedet 2004], ecto and endo cervical biopsies were mandated. Participants, colposcopists and pathologists were blinded to screening test results, ensuring that the colposcopy protocol would be followed independently of screening test results. Through this strategy, histology was available for diagnosis for more than 95% of participants undergoing colposcopy, irrespective of screening test results. As can be appreciated from Table 26, very few studies took such steps to ensure full blinding of outcome assessors (colposcopists and pathologists) and reached such a high proportion of histology ascertainment.

However, this approach to disease verification may have caused the discovery of more incipient or indolent lesions than would normally be unveiled by routine colposcopies. Liberal inclusion of all such lesions, while informative, yielded disease detection rates that are unlikely to reflect real-world community screening. Our conservative definition, based on disease confirmation in an excision specimen, reduced overdiagnosis bias. Indeed, misclassifying squamous metaplasia, a benign condition, for CIN can be the explanation for most LEEP-negative specimens [Ferenczy 1996, Stoler 2001]. We found that most high grade lesions that were HPV negative were not confirmed on the excision specimen, which suggests that HPV testing could play a role in pathology quality assurance [Castle 2007]. In summary, the conservative case definition, by limiting

misclassification of benign findings as CIN, provides more valid estimates of test performance.

More importantly, it is reassuring that the better net performance of HPV compared to Pap testing is unlikely to result in length-time bias (more favorable outcome due to identification of more indolent lesions), since disease was confirmed in the excision specimens slightly more often, albeit non-significantly, among HPV-positive than among Pap-positive high grade lesions diagnosed on biopsy (section 4.4.1).

5.2.5 Analysis

Our analysis strategy aimed to provide unbiased estimates of test performance and their precision. Correcting for verification bias allowed us to compute absolute rather than relative estimates. Absolute estimates reflect the true community-level screening performance that can inform cost-effectiveness studies. Conservative estimates reported in Table 14 are very similar, whichever case definition is used (crude, uncorrected or corrected). In the specific case of our study, when using the conservative case definition, the impact of correcting for verification bias is minimal for 2 reasons: (1) disease verification was very high and comparable among screen positive participants across study arms, and (2) no disease was found in screen negative participants. The fact that we did not find disease in screen negative participants (652 colposcopies) cannot guarantee that we would not have, had we investigated even more participants. However, no high grade disease was found in over 3000 exams reported in the literature [CCCaST, Schiffman 2000, Belinson 2001, Kulasingam 2002, Cuzick 2003, Petry 2003, Bigras

2005], which indicates that disease prevalence in this group of patients is probably exceedingly low.

The fact that corrected sensitivity estimates decreased dramatically when the liberal case definition was used may in fact be due to misclassification of disease status. Histological assessment of colposcopy-directed biopsies performs reasonably well in a high prevalence setting, for which it is usually reserved [Stoler 2001, Gage 2006]. However, the addition of a large number of screen negative participants to the case mix may have degraded colposcopy and biopsy performance [Schiffman 2007]. Only expert review will provide insights as to the reason for this discrepancy.

5.3 Public health implications

As was underlined in the introduction to this thesis, cervical cancer possesses all the characteristics that make screening relevant as a preventive strategy. Pap-based opportunistic screening and organized programs have contributed to the decline in incidence and mortality observed over the last decades in settings with appropriate quality control measures in place [Franco 2002]. However, the current strategy has limitations: incidence and mortality rates have leveled off or even increased in certain populations, making further gains unlikely [Davies 2005]. Moreover, the limited sensitivity of the currently used screening test contributes to the inefficiency of the process: it must be repeated frequently in order to reach acceptable programmatic sensitivity.

For this reason, many research initiatives have focused on the evaluation of new screening tests. Such technologies can be classified in 3 groups: direct visualization,

cellular tests (cytology), and viral tests (HPV testing). The evidence on visualization techniques underlines their lack of specificity, to a level that would be unacceptable in settings where more specific tests are available [Basu 2003, Claeys 2003, Cronje 2003, Gaffikin 2003, Sankaranarayanan 2003, Wright 2003b, Doh 2005, Goldie 2005, Mahe 2005, Sangwa-Lugoma 2006,]. However, these simple and inexpensive methods may play a crucial role in cervical cancer prevention in low- resource settings where no alternative is available. Ongoing large RCTs have even pointed to an impact on cervical cancer mortality by these methods [Sankaranarayanan 2007]. The available cellular tests, conventional cytology and LBC, most likely have similar performance characteristics and simply switching from one cellular test to another is unlikely to significantly improve cervical cancer screening [McCrory 1999, Broadstock, 2000, Hanselaar 2002, Moss 2003, NICE 2003, Noorani 2003, Karnon, 2004, Davey 2006, Hulstaert 2006].

Viral tests, either HC2 technology or standardized PCRs, offer the best hope of improving cervical cancer screening. The higher sensitivity, the better reproducibility and the more "upstream" focus on cervical carcinogenesis conferred by HPV testing, relatively to Pap cytology, would permit safely extending screening intervals, offsetting the costs incurred by increased colposcopy referrals on initial screen [Clavel 2004, Cuzick 2006a]. Frequent Pap testing has achieved good results in settings where quality assurance exists for screening, diagnostic, and therapeutic procedures. In such settings, it is difficult to predict if a change from Pap to HPV testing will reduce cervical cancer mortality even further. However, reliance on HPV testing may **improve efficiency.** There is evidence that a negative HPV test may offer the same level of protection over four to five years as the Pap test offers for two years [Clavel 2004, Davies 2005, Silins 2005, Bulkmans 2005].

More importantly, for women who get screened less frequently than recommended, a more sensitive test may prove to be very important.

Switching from cellular to viral tests will present many challenges. Given their performance characteristics, viral tests are not indicated in young women. Screening algorithms will need to be tailored to age, and possibly to other patient characteristics, such as markers of sexual activity. Before implementation is considered, clinical management algorithms will also need to be clarified and made available. **Translational research** will need to explore strategies which will assist in timely diffusion, and adherence to, new screening guidelines and related clinical management of screen positive women.

As with any change in technology, initial costs will undoubtedly be substantial. However, in-depth **cost analysis** may prove that, on the long run, those initial costs can be compensated by a decrease in the costs associated with invasive cervical cancer care and by a decrease in the number of necessary screening rounds. Consideration will also need to be given to the existing workforce of highly trained cytotechnicians. In many provinces a large proportion of this workforce is approaching retirement age and simple attrition may resolve the problem.

As HPV testing is incorporated into primary screening, health care providers should avoid unduly alarming HPV-positive women. Proper **health education** is a challenge as our understanding of the natural history of HPV infection and CIN has evolved rapidly, making it difficult to provide clear and consistent information and as such creating

confusion and stress [McCaffery 2003, McCaffery 2004, McCaffery 2006, Waller 2006]. The challenge is to inform HPV positive women of their status without stigmatizing them or creating excessive anxiety, while at the same time underlining the need for follow-up [Wright 2003c].

Fortunately, research has helped identify key messages which inform women in a positive manner [Anhang 2004a, Anhang 2004b, Maissi 2004, Waller 2005, Bulkmans 2006]. In the course of our trial we found that participants readily accept HPV testing when proper information was available. More importantly, Bulkmans [2006] showed that the introduction of HPV testing to the regular screening program has not changed participation rates as was feared by some.

6 CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

This thesis summarizes the main findings from the first phase of the Canadian Cervical Cancer Screening Trial (CCCaST), the first randomized controlled trial designed to compare HPV testing to Pap cytology as standalone tests for primary screening of cervical cancer and its precursors. Methodological limitations of previous studies were avoided: we used a randomized design, we included participants in the context of routine cervical cancer screening activities, screening tests were performed in regular community settings, the gold standard (colposcopy and biopsy) was applied blindly of screening test results, and the design allowed for correction of verification bias. These methodological features lend strength to our findings.

Our results clearly demonstrate the superior sensitivity of HPV testing for cervical cancer precursor screening. As was discussed throughout this thesis, the lack of sensitivity of the current screening test (Pap cytology) is responsible for efficacy and efficiency shortcomings of the current screening strategy. In settings with quality screening in place, approximately half of women diagnosed with cervical cancer have been screened. Moreover, the lack of sensitivity prompts professional agencies to recommend frequent and numerous screening rounds. In this context, the use of a more sensitive test could have an important impact on cervical cancer prevention activities.

The exploratory analyses provide insights into possible performance of test combination.

Triage algorithms did not perform as well as HPV testing alone in the context of our trial where cytotechnicians evaluating Pap smears were blinded to HPV test results. Co-testing with Pap and HPV reached a sensitivity of 100% (compared to 97% for HPV alone) but would, if adopted, result in a substantial increase in referrals, putting the cost-effectiveness of this strategy in question.

Spectrum effect analysis of Pap and HPV testing highlighted the variation of Pap performance across laboratories. This was not the case for HPV testing. Age and the presence of a new sexual partner in the last year were identified as important parameters, information easily obtained in clinical practice, which influence the performance of screening tests for cervical cancer. We found HPV testing to be less specific than Pap (by 3%), which can translate into an increase in costs and unnecessary procedures. However, older women in long(er) term relationships may benefit form the higher sensitivity of Pap testing while not suffering from the consequences of a decrease in specificity.

As HPV testing is starting to be advocated as the primary and sole screening test, research in cervical cancer prevention will need to address in priority strategies to triage HPV positive women. It is possible that the 'HPV followed by Pap' strategy we mentioned above may perform better in real life settings where case load would be reduced and cytotechnicians would be made aware that slides are from HPV positive women [Franco 2006]. Other strategies are also under investigation. The knowledge that most HPV infections are transient and that only persistent ones increase the risk of cancer may be useful in the contest of screening. It may be possible to repeat HPV testing and refer only women repeatedly testing positive for diagnostic evaluation [Bulk 2007]. The impact of

such a strategy on cancer prevention and costs, compared to direct referral, remains to be assessed. The risk of progression to invasion varies with HPV type, with HPV-16 and HPV-18 carrying a higher risk [Khan 2005, Bilk 2006]. Type specific HPV tests suitable for use in clinical laboratory offer the possibility of tailoring follow-up to the specific type(s) found in the cervix [Meijer 2006]. Steps by which oncogenic HPV infections progress to invasive cancer are mediated by proteins. Molecular tests which identify these proteins could help to target intervention on high grade lesions most destined to progress, while permitting a more conservative approach for others Cuzick 2006, von Knebel-Doeberitz 2006]. As the various molecular tests become available for use in clinical settings, the most promising triage avenues should be compared in controlled trials in population undergoing routine screening.

The current screening strategy typically entails three visits to treat precursor lesions: one for screening, one for diagnosis and one for treatment. This scenario is particularly ill-suited for resource-poor settings where access to health care may be an important problem [Denny 2006]. Moreover the various resources necessary to carry cervical cytology, quality standardized HPV tests or other molecular markers under investigation may be difficult to implement in such settings [Sankaranarayanan 2004; Sankaranarayanan 2005]. Given that most cervical cancer cases are diagnosed in resource-poor settings [Parkin 2002], there is a sense of urgency to develop tailored screening measures. Future research will need to build on ongoing projects and compare the efficacy and cost-effectiveness of alternative scenarios [Mandelblatt 2002, Goldie 2005]. A one visit scenario where rapid and simple HPV tests (under development,

reviewed in Cuzick 2006) followed by visual inspection of positive women and immediate treatment of identified lesion could should be evaluated.

The success of HPV vaccines opens a new era for cervical cancer prevention. However, there are still many unknowns such as long term safety and efficacy, duration of protection, best time to offer vaccination, optimal vaccine delivery mechanisms, etc. National and international associations have agreed on key issues that need to be clarified [Franco 2006b, Hidelsheim 2006]. For one, it is still unclear if and how screening should be tailored to the vaccination status of women. It is unlikely that universal health care programs will fund HPV vaccination programs indefinitely if no gains (in terms of cost and/or resource utilization) can be made by scaling back screening activities. For those who are vaccinated, continued screening with HPV testing may provide the added benefit of HPV surveillance [Franco 2006a]. If, as we expect, vaccines are successful in substantially reducing the number of high grade cervical lesion, cytology and colposcopy performance may suffer [Schiffman 2007]. These changes will need to be assessed and alternative management options proposed [Jeronimo 2006].

Screening with HPV tests and vaccination are recent alternatives to cervical cancer prevention and will require research to identify the most efficacious and cost-effective options to reduce mortality from cervical cancer. But while we search for the optimal strategy, screening with a more sensitive test offers the hope of further improving cervical cancer prevention.

BIBLIOGRAPHY

Abraido-Lanza AF, Chao MT, Gammon MD. Breast and cervical cancer screening among latinas and non-latina whites. American Journal of Public Health 2004; 94: 1393-1398.

Acharya G, Kjeldberg I, Hansen SM et al. Pregnancy outcome after loop electrosurgical excision procedure for the management of cervical intraepithelial neoplasia. Archives of Gynecology and Obstetrics 2005; 272: 109-112.

Adami HO, Baron JA, Rothman KJ. Ethics of a prostate Cancer screening trial. The Lancet. 1994:343:958-60.

Agorastos T, Dinas K, Lloveras B et al. Human papillomavirus testing for primary screening in women at low risk of developing cervical cancer. The Greek experience. Gynecologic Oncology 2005; 96: 714-720.

Akaike H. On entropy maximization principle, in: P.R. Krishnaiah ed., Applications of statistics (North-Holland: Amsterdam). 1977.

Altman DG, Bland JM. Diagnostic tests 2: Predictive values. British Medical Journal 1994a; 309: 102.

Altman DG, Bland JM. Diagnostic tests 3: Receiver operating characteristic plots. British Medical Journal 1994b; 309: 188-189.

Anhang R, Goodman A, Goldie SJ. HPV communication: Review of existing research and recommendations for patient education. CA: a Cancer Journal for Clinicians 2004a; 54: 248-259.

Anhang R, Nelson JA, Telerant R et al. Acceptability of self-collection of specimens for HPV DNA testing in an urban population. Journal of Women's Health 2005; 14: 721-728.

Anhang R, Wright TC Jr, Smock L et al. Women's desired information about human papillomavirus. Cancer 2004b; 100: 315-320.

Anttila A, Pukkala E, Soderman B et al. Effect of organized screening on cervical cancer incidence and mortality in Finland, 1963-1995: recent increase in cervical cancer incidence. International Journal of Cancer 1999; 83: 59-65.

Anttila A, Ronco G, Clifford G et al. Cervical cancer screening programmes and policies in 18 European countries. British Journal of Cancer 2004a; 91: 935-941.

Arbyn M, Buntinx F, Van Ranst M et al. Virologic versus cytologic triage of women with equivocal Pap smears: a meta-analysis of the accuracy to detect high-grade intraepithelial neoplasia. Journal National Cancer Institute Monographs 2004; 96: 280-293.

Arbyn M, Sasieni P, Meijer CJ et al. Chapter 9: Clinical applications of HPV testing: A summary of meta-analyses. Vaccine. 2006 Aug 21; 24 Suppl 3:S78-89.

Armitage P, Berry G. Diagnostic tests and screening procedures. In: Statistical Methods in Medical Research. 3rd ed. Blackwell Scientific Publications. Oxford, England, 1994, pp 522-529.

ALTS (ASCUS-LSIL Triage Study) Group. Results of a randomized trial on the management of cytology interpretations of atypical squamous cells of undetermined significance. American Journal of Obstetric and Gynecology 2003; 188: 1383-1392.

Austin RM, Ramzy I. Increased detection of epithelial cell abnormalities by liquid-based gynecologic cytology preparations. A review of accumulated data. Acta Cytologica 1998; 42: 178-184.

Austoker J. Gaining informed consent for screening. British Medical Journal 1999; 319: 722-723.

Bailar JC. Introduction In: Shapiro SH, Louis TA, eds. Clinical trials issues and approaches. Marcel Dekker Inc., New York, USA, 1983,pp 1-12.

Bais AG, Reolj M, Snijders PJ et al. Triage using HPV-testing in persistent borderline and mildly dyskaryotic smears: proposal for new guidelines. International Journal of Cancer 2005; 116: 122-129.

Bais AG, van Kemenade FJ, Berkhof J et al. Human papillomavirus testing on self-sampled cervicovaginal brushes: An effective alternative to protect nonresponders in cervical screening programs. International Journal of Cancer 2007; 120: 1505-1510.

Baldwin P. Laskey R. Coleman N. Translational approaches to improving cervical screening. Nature Reviews - Cancer 2003; 3: 217-226.

Baldwin S, Santos C, Brown EM et al. Comparison of type-specific human papillomavirus data from self and clinician directed sampling. Gynecologic Oncology 2005; 97: 612-617.

Bastani R, Berman BA, Belin TR. Increasing cervical cancer screening among underserved women in a large urban county health system. Medical Care 2002; 40: 891-907.

Basu PS, Sankaranarayanan R, Mandal R et al. Visual inspection with acetic acid and cytology in the early detection of cervical neoplasia in Kolkata, India. International Journal of Gynecological Cancer 2003; 13: 626-632.

Bauer HM, Hildesheim A, Schiffman MH et al. Determinants of genital human papillomavirus infection in low-risk women in Portland, Oregon. Sexually Transmitted Diseases 1993; 20: 274-278.

Begg CB. Advances in statistical methodology for diagnostic medicine in the 1980's. Statistics in Medicine 1991; 10: 1887-1895.

Begg CB. Biases in the assessment of diagnostic tests. Statistics in Medicine 1987; 6: 411-423.

Belinson J, Qiao YL, Pretorius R et al. Shanxi Province cervical cancer screening study: a cross-sectional comparative trial of multiple techniques to detect cervical neoplasia. Gynecologic Oncology 2001; 83: 439-444.

Belinson JL, Qiao YL, Pretorius R et al. Shanxi Province cervical cancer screening study II: Self-sampling for high-risk human papillomavirus compared to direct sampling for human papillomavirus and liquid based cervical cytology. International Journal of Gynecological Cancer 2003; 13: 819-826.

Benedet JL, Matisic JP. Bertrand MA. The quality of community colposcopic practice. Obstetrics & Gynecology 2004; 103: 92-100.

Berkhof J, de Bruijne MC, Zielinski GD et al. Evaluation of cervical screening strategies with adjunct high-risk human papillomavirus testing for women with borderline or mild dyskaryosis. International Journal of Cancer 2006; 118: 1759-1768.

Bigras G, De Marval F. The probability for a Pap test to be abnormal is directly proportional to HPV viral load: Results from a Swiss study comparing HPV testing and liquid-based cytology to detect cervical cancer precursors in 13,842 women. British Journal of Cancer 2005; 93: 575-581.

Blumenthal PD, Gaffikin L, Chirenje ZM et al. Adjunctive testing for cervical cancer in low resource settings with visual inspection, HPV, and the Pap smear. International Journal of Gynecology and Obstetrics 2001; 72: 47-53.

Bosch FX, Harper D. Prevention strategies of cervical cancer in the HPV vaccine era. Gynecologic Oncology 2006; 103: 21-24.

Bosch FX, Lorincz A, Munoz N et al. The causal relation between human papillomavirus and cervical cancer. Journal of Clinical Pathology 2002; 55: 244-265.

Bosch FX, Manos MM, Munoz N et al. Prevalence of human papillomavirus in cervical cancer - a worldwide perspective. Journal National Cancer Institute Monographs 1995; 87: 796-802.

Bossuyt PM, Reitsma JB, Bruns DE et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Annals of Clinical Biochemistry 2003: 40: 357-363.

Breen N, Meissner HI. Toward a system of cancer screening in the United States: Trends and opportunities. Annual Review of Public Health 2005; 226: 561-582.

Brinton LA. Ways that women may possibly reduce their risk of breast cancer. Journal of National Cancer Institute 1994; 86: 371-372.

Broadstock M. Effectiveness and costs effectiveness of automated and semi-automated cervical screening devices: a systematic review. New Zealand Technology Assessment Report (NZHTA), 2000; NZHTA Report Volume 3 Number 1.

Bulk S, Berkhof J, Bulkmans NW *et al.*, Preferential risk of HPV16 for squamous cell carcinoma and of HPV18 for adenocarcinoma of the cervix compared to women with normal cytology in The Netherlands, British journal of Cancer 2006; 94: 171–175.

Bulk S, Bulkmans NWJ, Berkhof J et al. Risk of high-grade cervical intra-epithelial neoplasia based on cytology and high-risk HPV testing at baseline and at 6-months. International Journal of Cancer 2007; 121: 361-367.

Bulkmans NW, Rozendaal L, Voorhorst FJ et al. Long-term protective effect of high-risk human papillomavirus testing in population-based cervical screening. British Journal of Cancer 2005; 92: 1800-1802.

Bulkmans NW, Bulk S, Ottevanger MS et al. Implementation of human papillomavirus testing in cervical screening without a concomitant decrease in participation rate. Journal of Clinical Pathology 2006; 59: 1218-1220.

Calman, KC. The ethics of allocation of scarce health resources. Journal of Medical Ethics 1994; 20: 71-74.

Castellsague X, Munoz N. Cofactors in human papillomavirus carcinogenesis - Role of parity, oral contraceptives, and tobacco smoking. Journal National Cancer Institute Monographs 2003; 31: 20-28.

Castle PE, Jeronimo J, Schiffman M et al. Age-related changes of the cervix influence human papillomavirus type distribution. Cancer Research 2006; 66: 1218-1224.

Castle PE, Schiffman M, Herrero R et al. A prospective study of age trends in cervical human papillomavirus acquisition and persistence in Guanacaste, Costa Rica. Journal of Infectious Diseases 2005; 191:1808-1816.

Castle PE, Stoler MH, Solomon D et al. The relationship of community biopsy-diagnosed cervical intraepithelial neoplasia grade 2 to the quality control pathology-reviewed diagnoses - An ALTS report. American Journal of Clinical Pathology 2007; 127: 805-815.

CCS. Canadian Cancer Society/National Cancer institute of Canada: Canadian Cancer statistics 2007. Toronto, Canada, 2007. April 2007, ISSN 0835-2976. Accessed July 2007 at http://129.33.170.32/vgn/images/portal/cit_86751114/36/15/1816216925cw_2007stats_en.pdf

Chua KL, Hjerpe A. Persistence of human papillomavirus (HPV) infections preceding cervical carcinoma. Cancer 1996; 77: 121-127.

CIOMS. Council for International Organizations of Medical Sciences. International Ethical Guidelines for Biomedical Research Involving Human Subjects 2002. ISBN 92-9036-075-5. Accessed July 2007 at http://www.cioms.ch/frame_guidelines_nov_2002.htm.

Claeys P, De Vuyst H, Gonzalez C et al. Performance of the acetic acid test when used in field conditions as a screening test for cervical cancer. Tropical Medicine & International Health 2003; 8: 704-709.

Clavel C, Cucherousset J, Lorenzato M et al. Negative human papillomavirus testing in normal smears selects a population at low risk for developing high-grade cervical lesions. British Journal of Cancer 2004; 90: 1803-1808.

Clavel C, Masure L, Bory JP et al. Human papillomavirus testing in primary screening for the detection of high-grade cervical lesions: a study of 7932 women. British Journal of Cancer 2001; 84: 1616-1623.

Coldman A. HPV Testing for Cervical Cancer Screening Study. Registered at clinicalTrials.gov at http://clinicaltrials.gov/ct/show/NCT00461760;jsessionid=AA49406A10C3-11A16E38BF84343BC8BF?order=1.

Collins S, Mazloomzadeh S, Winter H et al. High incidence of cervical human papillomavirus infection in women during their first sexual relationship. British Journal of Obstetrics and Gynecology 2002; 109: 96-98.

Costa S, De Nuzzo M, Rubino A et al. Independent determinants of inaccuracy of colposcopically directed punch biopsy of the cervix. Gynecologic Oncology 2003; 90: 57-63.

Coste J, Cochand-Priollet A, de Cremoux P et al. Cross sectional study of conventional cervical smear, monolayer cytology, and human papillomavirus DNA testing for cervical cancer screening. British Medical Journal 2003; 326: 733A-736A.

Coughlin SS, Trock B, Criqui MH et al. The logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test. Journal of Clinical Epidemiology 1992; 45: 1-7.

Cox JT, Lornizc AT, Schiffman MH et al. Human paillomavirus testing by hybrid capture appears to be useful in triaging women with a cytologic diagnosis of atypical squamous cells of underminate significance. Obstetrics and Gynecology 1995; 172: 946-954.

Cronin KA, Weed DL, Connor RJ et al. Case-control studies of cancer screening: theory and practice. Journal of the National Cancer Institute 1998; 90: 498-504.

Cronje HS, Parham GP, Cooreman BE et al. A comparison of four screening methods for cervical neoplasia in a developing country. American Journal of Obstetric and Gynecology 2003; 188: 395-400.

Cuzick J, Beverley E, Ho L et al. HPV testing in primary screening of older women. British Journal of Cancer 1999a; 81: 554-558.

Cuzick J, Clavel C, Petry KU et al. Overview of the European and North American studies on HPV testing in primary cervical cancer screening. International Journal of Cancer 2006a; 119: 1095-1101.

Cuzick J, Mayrand MH, Ronco G et al. New dimensions in cervical cancer screening. Vaccine 2006b; 24(Suppl 3): 90-97.

Cuzick J, Sasieni P, Davies P et al. A systematic review of the role of human papillomavirus testing within a cervical screening programme. Health Technology Assessment 1999b; 3 (14).

Cuzick J, Szarewski A, Cubie H et al. Management of women who test positive for high-risk types of human papillomavirus: the HART study. The Lancet 2003; 362: 1871-1876.

Cuzick J, Szarewski A, Terry G et al. Human papillomavirus testing in primary cervical screening. The Lancet 1995; 345: 1533-1536.

Cyrus-David MS, Michielutte R, Paskett ED et al. Cervical cancer risk as a predictor of Pap smear use in rural North Carolina. Journal of Rural Health 2002; 18: 67-76.

Dannecker C, Siebert U, Thaler CJ et al. Primary cervical cancer screening by self-sampling of human papillomavirus DNA in internal medicine outpatient clinics. Annals of Oncology 2004; 15: 863-869.

Davey DD, Nielsen M, Souers R et al. Pap smear specimen adequacy in laboratories surveyed by the CAP. Modern Pathology 1996; 9: 34A.

Davey E, Barrat A, Irwig L et al. Effect of study design and quality on unsatisfactory rates, cytology classifications, and accuracy in liquid-based versus conventional cervical cytology: a systematic review. The Lancet 2006; 367: 122-132.

Davies P, Arbyn M, Dillner J et al. A report on the current status of European research on the use of human papillomavirus testing for primary cervical cancer screening. International Journal of Cancer 2005; 118: 791-796.

de Cremoux P, Coste J, Sastre-Garau X et al. Efficiency of the hybrid capture 2 HPV DNA test in cervical cancer screening - A study by the French society of clinical cytology. American Journal of Clinical Pathology 2003; 120: 492-499.

Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. British Medical Journal 2004; 329: 168-169.

Denny L, Sankaranarayanan R. Chapter 8: Screening for cervical cancer in developing countries. Vaccine 2006; 24 suppl 3: S71-S77.

Doh AS, Nkele NN, Achu P et al. Visual inspection with acetic acid and cytology as screening methods for cervical lesions in Cameroon. International Journal of Gynecological and Obstetrics 2005; 89: 167-173.

Dunne EF, Unger ER, Sternberg M et al. Prevalence of HPV infection among females in the United States. Journal of the American Medical Association 2007; 297: 813-819.

Dzuba IG, Diaz EY, Allen B et al. The acceptability of self-collected samples for HPV testing vs. the Pap test as alternatives in cervical cancer screening. Journal of Women's Health & Gender-Based Medicine 2002; 11: 265-274.

Elfgren K, Rylander E, Radberg T et al. Colposcopic and histopathologic evaluation of women participating in population-based screening for human papillomavirus deoxyribonucleic acid persistence. American Journal of Obstetrics and Gynecology 2005; 193: 650-657.

Ferenczy A, Choukroun D, Arseneau J. Loop electrosurgical excision procedure for squamous intraepithelial lesions of the cervix: Advantages and potential pitfalls. Obstetrics and Gynecology 1996; 87: 332-337.

Fletcher RH, Fletcher SW, Wagner EH. Prevention. In: Satterfield TS, ed. Clinical epidemiology the essentials. 3rd edition. Williams and Wilkins, Baltimore, USA, 1996, pp 168-195.

Forbes C, Jepson R, Martin-Hirsch P. Interventions targeted at women to encourage the uptake of cervical screening. Cochrane Gynaecological Cancer Group, Cochrane Database of Systematic Reviews 2007; 2.

Forrest S, McCaffery K, Waller J et al. Attitudes to self-sampling for HPV among Indian, Pakistani, African-Caribbean and white British women in Manchester, UK. Journal of Medical Screening 2004; 11: 85-88.

Franco EL. Chapter 13: Primary screening of cervical cancer with human papillomavirus tests. Journal National Cancer Institute Monographs 2003; 31: 89-96.

Franco EL. Statistical issues in human papillomavirus testing and screening. Clinics in Laboratory Medicine 2000; 20: 345-367.

Franco EL, Bosch FX, Cuzick J et al. Chapter 29: Knowledge gaps and priorities for research on prevention of HPV infection and cervical cancer. Vaccine 2006b ;24 Suppl 3:S242-S249.

Franco EL, Cuzick J, Hildesheim A et al. Chapter 20: Issues in planning cervical cancer screening in the era of HPV vaccination. Vaccine. 2006a; 24 Suppl 3:S171-S177.

Franco EL, Ferenczy A. Assessing gains in diagnostic utility when human papillomavirus testing is used as an adjunct to Papanicolaou smear in the triage of women with cervical cytologic abnormalities. American Journal of Obstetrics and Gynecology 1999b; 181: 382-386.

Franco EL, Harper DM. Vaccination against human papillomavirus infection: A new paradigm in cervical cancer control. Vaccine 2005; 23: 2388-2394.

Franco EL, Rohan TE. Evidence-based policy recommendations on screening and prevention. Cancer Precursors: Epidemiology, Detection, and Prevention. Springer-Verlag, New York, USA, 2002, pp 389-403.

Franco EL, Rohan TE, Villa LL. Epidemiologic evidence and human papillomavirus infection as a necessary cause of cervical cancer. Journal National Cancer Institute Monographs 1999a; 91: 506-511.

Freedman B. Equipoise and the ethics of clinical research. New England Journal of Medicine 1987; 317:141-145.

The FUTURE II Study Group. Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. New England Journal of Medicine 2007; 356: 1915-1927.

Gaffikin L, Lauterbach M, Blumenthal PD. Performance of visual inspection with acetic acid for cervical cancer screening: A qualitative summary of evidence to date. Obstetrical & Gynecological Survey 2003; 58: 543-550.

Gage JC, Hanson VW, Abbey K et al. Number of cervical biopsies and sensitivity of colposcopy. ASCUS LSIL Triage Study (ALTS) Group. Obstetrics and Gynecology 2006; 108: 264-272.

Garland SM, Hernandez-Avila M, Wheeler CM et al. Quadrivalent vaccine against human papillomavirus to prevent anogenital diseases. New England Journal of Medicine 2007; 356: 1928-1943.

Goldie SJ, Gaffikin L, Goldhaber-Fiebert JD et al. Cost-effectiveness of cervical-cancer screening in five developing countries. New England Journal of Medicine 2005; 353: 2158-2168.

Goldie SJ, Kohli M, Grima D et al. Projected clinical benefits and cost-effectiveness of a human papillomavirus 16/18 vaccine. Journal of the National Cancer Institute Monographs 2004; 96: 604-615.

Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses American Journal of Epidemiology 1995; 142: 1255-1264.

Greenwald P, Costlow R, Prorok P et al. Screening in the context of cancer control. In: Screening for Cancer. AB Miller, editor. Academic Press, Orlando, USA, 1985, pp 25-48.

Guido R, Schiffman M, Solomon D et al. Postcolposcopy management strategies for women referred with low-grade squamous intraepithelial lesions or human papillomavirus DNA-positive atypical squamous cells of undetermined significance: a two-year prospective study. American Journal of Obstetrics and Gynecology 2003; 188: 1401-1405.

Hall S, Lornizc AT, Shah F et al. Human papilloma DNA detection in cervical specimens by hybrid capture: correlation with cytologic and histologic diagnoses of squamous intraepithelial lesions of the cervix. Gynecologic Oncology 1996; 62: 535-539.

Hanselaar AGJM. Criteria for organized cervical screening programs: Special emphasis on the Netherlands program. Acta Cytologica 2002; 46: 619-629.

Harper DM, Franco EL, Wheeler CM et al. Efficacy of a bivalent L1 virus-like particle vaccine in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomised controlled trial. The Lancet 2004; 364: 1757-1765.

Harper DM, Franco EL, Wheeler CM et al. Sustained efficacy up to 4.5 years of a bivalent L1 virus-like particle vaccine against human papillomavirus types 16 and 18: follow-up from a randomised control trial. The Lancet 2006; 367: 1247-1255.

Harper R, Reeves B. Reporting of precision of estimates for diagnostic accuracy: a review. British Medical Journal 1999; 318: 1322-1323.

Hennekens CH, Buring JE. Screening. In: Mayrent SL. ed. Epidemiology in medicine. Little Brown & Co., Boston, USA, 1987, pp 327-350.

Hildesheim A, Bratti MC, Edwards RP et al. Collection of cervical secretions does not adversely affect Pap smears taken immediately afterward. Clinical and Diagnostic Laboratory Immunology 1998; 5: 491-493.

Hildesheim A, Hadjimichael O, Schwartz PE et al. Risk factors for rapid-onset cervical cancer. American Journal of Obstetrics and Gynecology 1999; 180: 571-577.

Hildesheim A, Herrero R, Castle PE et al. HPV co-factors related to the development of cervical cancer: results from a population-based study in Costa Rica. British Journal of Cancer 2001; 84: 1219-1226.

Hildesheim A, Schiffman MJ, Gravitt PE et al. Persistence of type-specific human papillomavirus infection among cytologically normal women. Journal of Infectious Diseases 1994; 169: 235-240.

Ho GY, Kadish AS, Burk RD et al. HPV 16 and cigarette smoking as risk factors for high-grade cervical intra-epithelial neoplasia. International Journal of Cancer 1998a; 78: 281-285.

Ho GYF, Bierman R, Beardsley L et al. Natural history of cervicovaginal papillomavirus infection in young women. New England Journal of Medicine 1998b; 338: 423-428.

Ho GYF, Burk RD, Klein S et al. Persistent genital human papillomavirus infection as a risk factor for persistent cervical dysplasia. Journal of the National Cancer Institute Monographs 1995; 87: 1365-1371.

Hopman EH, Voorhorst FJ, Kenemans P et al. Observer agreement on interpreting colposcopic images of CIN. Gynecologic Oncology 1995; 58: 206–209.

Howlett R. Elements of organized cancer screening. Ontario screening program, Cancer care Ontario 2006. Accessed July 2007at http://www.osmt.org/Elements%20of%20Cancer%20Screening.pdf.

Hulstaert F, Arbyn M, Huybrechts M et al. Cervical cancer screening and human papillomavirus (HPV) testing. Health technology assessment (HTA). Brussels, Belgium, Belgian Health Care Knowledge Centre (KCE), Report 38C (D/2006/10.273/37), 2006.

IARC. IARC Monographs on the evaluation of carcinogenic risks to humans. Vol 64:.Human papillomaviruses. IARC press 1995, Lyon, France; 409 pages.

IARC. Handbooks of cancer prevention, vol. 10. Cervix cancer screening. IARC press 2005 Lyon, France.

Jepson RG, Forbes CA, Sowden AJ et al. Increasing informed uptake and non-uptake of screening: evidence from a systematic review. Health Expectations: An International Journal of Public Participation in Health Care and Health Policy 2001; 4: 116-126.

Jeronimo J, Schiffman M. Colposcopy at a crossroads. American Journal of Obstetrics and Gynecology 2006; 195: 349-353.

Johnson K, Canadian Task Force. Screening for human papillomavirus infection in asymptomatic women. Update: 1. Canadian Medical Association Journal 1995; 152: 489-493.

Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: Defining priorities to reduce cancer disparities in different geographic regions of the world. Journal of Clinical Oncology 2006; 24: 2137-2150.

Karnon J, Peters J, Platt J et al. Liquid-based cytology in cervical screening: an updated rapid and systematic review and economic analysis. Health Technology Assessment 2004; 8.

Khan MJ, Castle PE, Lorincz AT et al. The elevated 10-year risk of cervical precancer and cancer in women with human papillomavirus (HPV) type 16 or 18 and the possible utility of type-specific HPV testing in clinical practice. Journal National Cancer Institute Monographs 2005; 97: 1072-1079.

Kim JJ, Wright TC, Goldie SJ. Chapter 13: Cost-effectiveness of alternative triage strategies for atypical squamous cells of undetermined significance. Journal of the American Medical Association 2002; 287: 2382-2390.

Kjaer SK, Van den Brule AJ, Paull G et al. Type specific persistence of high risk human papillomavirus (HPV) as indicator of high grade cervical squamous intraepithelial lesions in young women: population based prospective follow up study. British Medical Journal 2002; 325: 572-579.

Kolstad P, Klem V. Long-term follow-up of 1121 cases of carcinoma in situ. Obstetrics and Gynecology 1976; 48: 125-129.

Kotaniemi-Talonen L, Nieminen P, Anttila A et al. Routine cervical screening with primary HPV testing and cytology triage protocol in a randomised setting. British Journal of Cancer 2005; 93: 862-867.

Koutsky LA, Harper DM. Chapter 13: Current findings from prophylactic HPV vaccine trials. Vaccine. 2006;24: S114-S121.

Koutsky LA, Holmes KK, Critchlow CW. A cohort study of the risk of cervical intraepithelial neoplasia grade 2 or 3 in relation to papillomavirus infection. New England Journal of Medicine 1992; 327:1 272-1278.

Kuhn L, Denny L, Pollack A et al. Human papillomavirus DNA testing for cervical cancer screening in low-resource settings. Journal National Cancer Institute Monographs 2000; 92: 818-825.

Kulasingam SL, Hughes JP, Kiviat NB et al. Evaluation of human papillomavirus testing in primary screening for cervical abnormalities - Comparison of sensitivity, specificity, and frequency of referral. Journal of the American Medical Association 2002; 288: 1749-1757.

Kulasingam SL, Kim JJ, Lawrence WF et al. Cost-effectiveness analysis based on the atypical squamous cells of undetermined significance/low-grade squamous intraepithelial lesions triage study (ALTS). Journal of the National Cancer Institute 2006; 98: 92–100.

Kulasingam SL, Myers ER. Potential health and economic impact of adding a human papillomavirus vaccine to screening programs. Journal of the American Medical Association 2003; 290: 781-789.

Kyrgiou M, Koliopoulos G, Martin-Hirsch P et al. Obstetric outcomes after conservative treatment for intraepithelial or early invasive cervical lesions: systematic review and meta-analysis. The Lancet 2006; 367: 489-498.

Lebacqz K. Ethical aspects of clinical trials. In: Shapiro SH, Louis TA, eds. Clinical trials issues and approaches. Marcel Dekker Inc., New York, USA, 1983, pp 81-98.

Lee JM. Screening and Informed Consent. New England Journal of Medicine 1993; 328: 438-440.

Leyden WA, Manos MM, Geiger AM et al. Cervical cancer in women with comprehensive health care access: attributable factors in the screening process. Journal of the National Cancer Institute 2005; 97: 675-83.

Lippman A, Melnychuk R, Shimmin C, et al., Human papillomavirus, vaccines and women's health: questions and cautions. CMAJ 2007; 177:484-7.

Liu S, Semenciw R, Probert A et al. Cervical cancer in Canada: changing patterns in incidence and mortality. International Journal of Gynecology and Cancer 2001; 11: 24–31.

Loizzi P, Carriero C, Di Gesu A et al. Rational use of cryosurgery and cold knife conization for treatment of cervical intraepithelial neoplasia. European Journal of Gynaecological Oncology 1992; 13: 507-513.

Mahe C, Gafikkin L. Screening test accuracy studies: how valid are our conclusions? Application to visual inspection methods for cervical screening. Cancer Causes & Control 2005; 16: 657-666.

Mahon J. Direct and indirect randomized trials of screening: the A's and D's of evidence-based clinical practice guidelines. Canadian Medical Association Journal 2000; 162: 1002-1004.

Maissi E, Marteau TM, Hankins M et al. Psychological impact of human papillomavirus testing in women with borderline or mildly dyskaryotic cervical smear test results: cross sectional questionnaire study. British Medical Journal 2004; 328: 1293-1298.

Mandelblatt JS, Lawrence WF, Gaffikin L, et al. Costs and benefits of different strategies to screen for cervical cancer in less-developed countries. Journal of the National Cancer Institute 2002; 94: 1469-1483.

Manhart LE, Koutsky LA. Do condoms prevent genital HPV infection, external genital warts, or cervical neoplasia? Sexually Transmitted Diseases 2002; 29: 725-735.

Manos MM, Kinney WK, Hurley LB et al. Identifying women with cervical neoplasia - Using human papillomavirus DNA testing for equivocal Papanicolaou results. Journal of the American Medical Association 1999; 281: 1605-1610.

Martin-Hirsch PL, Paraskevaidis E, Kitchener H. Surgery for cervical intraepithelial neoplasia. Cochrane Database of Systematic Reviews 2 2005.

Massad LS, Collins YC. Strength of correlations between colposcopic impression and biopsy histology. Gynecologic Oncology 2003; 89: 424-428.

Maticka-Tyndale E. Sexual health and Canadian youth: How do we measure up? The Canadian Journal of Human Sexuality 2001; 10: 1-17.

Mayrand MH, Duarte-Franco E, Coutlée F et al. for the CCCaST Study Group. Randomized study of Human papillomavirus testing versus Pap cytology in the primary screening for cervical cancer precursors: Design, methods and preliminary accrual results of The Canadian Cervical Cancer Screening Trial (CCCaST). Internationa Journal of Cancer 2006; 119: 615-623.

Mayrand MH, Duarte-Franco E, Rodrigues I et al. for the CCCaST Study Group. Randomized controlled trial of human papillomavirus testing versus Pap cytology for primary screening of cervical cancer precursors: results from the Canadian Cervical Cancer Screening Trial (CCCaST). New England Journal of Medicine 2007; accepted for publication.

McCaffery K, Forrest S, Waller J et al. Attitudes towards HPV testing: a qualitative study of beliefs among Indian, Pakistani, African-Caribbean and white British women in the UK. British Journal of Cancer 2003; 88: 42-46.

McCaffery K, Waller J, Forrest S et al. Testing positive for human papillomavirus in routine cervical screening: examination of psychosocial impact. British Journal of Obstetrics and Gynecology 2004; 111: 1437-1443.

McCaffery K, Waller J, Nazroo J et al. Social and psychological impact of HPV testing in cervical screening: a qualitative study. Sexually Transmitted Infections 2006; 82: 169-174.

McCrory DC, Matchar DB, Bastian L et al. Evaluation of cervical cytology. Evidence Report/Technology Assessment No. 5. Agency for Health Care Policy and Research (AHCPR) 1999; 99-E010.

McKneally MF, McKneally BM, Dickens EM et al. Bioethics for Clinicians: Resource Allocation. Canadian Medical Association Journal 1997; 157: 163-7.

Meijer CJ, Snijders PJ, Castle PE. Clinical utility of HPV genotyping. Gynecologic Oncology 2006; 103: 12-17.

Miles A, Cockburn J, Smith RA et al. A perspective from countries using organized screening programs. Cancer 2004; 101 (Suppl 5): 1201-1213.

Miller AB. Design of cancer screening trials/randomized trials for evaluation of cancer screening. World Journal of Surgery 2006; 30: 1152-1156.

Miller AB. Principles of screening and of evaluation of screening programs. In: Screening for Cancer. AB Miller, ed. Academic Press, Orlando, USA, 1985, pp 3-25.

Miller AB. The (in)efficiency of cervical screening in Europe. European Journal of Cancer 2002; 38: 321–326.

Miller AB, Anderson G, Brisson J et al. Report of a national workshop on screening for cancer of the cervix. Canadian Medical Association Journal 1991; 145: 1301-1324.

Miller BA, Kolonel LN, Bernstein L. Racial/ethnic patterns of cancer in the United States 1988-1992. National Cancer Institute, Bethesda, USA, 1994, NIH Publications no. 96-4101.

Mitchell MF, Schottenfeld D, Tortolero-Luna G et al. Colposcopy for the diagnosis of squamous intraepithelial lesions: a meta-analysis. Obstetrics and Gynecology 1998a; 91: 626–631.

Mitchell MF, Tortolero-Luna G, Cook E et al. A randomized trial of cryotherapy, laser vaporization, and loop electrosurgical excision for treatment of squamous intraepithelial lesions of the cervix. Obstetrics and Gynecology 1998b; 92: 737-744.

Moher D, Schulz KF, Altman DG et al. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel group randomized trials. The Lancet 2001; 357: 1191–1194.

Molano M, Van den Brule A, Plummer M et al. Determinants of clearance of human papillomavirus infections in Colombian women with normal cytology: a population-based, 5-year follow-up study, American Journal of Epidemiology 2003; 158: 486–494.

Morrison AS. Cancer of the cervix. In: Kelsey JL, Marmot MG, Stolley PD, Vessey MP, eds. Screening in chronic disease. 2nd edition. Oxford University Press, New York, USA, 1992, pp 216-234.

Moscicki AB, Shiboski S, Hills NK et al. Regression of low-grade squamous intra-epithelial lesions in young women. The Lancet 2004; 364: 1678–1683.

Moscicki AB, Shiboski S, Broering J et al. The natural history of human papillomavirus infection as measured by repeated DNA testing in adolescent and young women. Journal of Pediatrics 1998; 132: 277–284.

Moscicki AB, Hills N, Shiboski S et al. Risks for incident human papillomavirus infection and low-grade squamous intraepithelial lesion development in young females. Journal of the American Medical Association 2001; 285: 2995–3002.

Moss SM, Gray A, Legood R et al. Evaluation of HPV/LBC cervical screening pilot studies: first report to the Department of Health on evaluation of LBC. Surrey Institute of Cancer Research 2003.

Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. Annals of Internal Medicine 2002; 137: 598-602.

Munoz N, Bosch FX, De Sanjose S et al. The causal link between human papillomavirus and invasive cervical cancer: A population-based case-control study in Colombia and Spain. International Journal of Cancer 1992; 52: 743-749.

Munoz N, Bosch FX, De Sanjose S et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. New England Journal of Medicine 2003; 348: 518-527.

Munoz N, Bosch FX, Castellsague X et al. Against which human papillomavirus types shall we vaccinate and screen? The international perspective. International Journal of Cancer 2004; 111: 278-285.

NACI. National Advisory Committee on Immunization. Statement on Human Papillomavirus vaccine. CCDR 2007; 33: 1-32.

Nanda K, McCrory DC, Myers ER et al. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalitites: a systematic review. Annals of Internal Medicine 2000; 132: 810-819.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research.The Belmont Report: Ethical Principles and Guidelines for the protection of human subjects of research 1979. Accessed July 2007 at http://ohsr.od.nih.gov/guidelines/belmont.html.

NICE. National Institute for Clinical Excellence. Guidance on the use of liquid-based cytology for cervical screening-. Technology Appraisal Guidance 2003; 69 (296):31.

Nieminen P, Kallio M, Anttila A et al. Organised versus spontaneous Pap-smear screening for cervical cancer, a case control study. International Journal of Cancer 1999; 83: 55-58.

Nobbenhuis MAE, Helmerhorst TJM, Van den Brule AJC et al. Primary screening for high risk HPV by home obtained cervicovaginal lavage is an alternative screening tool for unscreened women. Journal of Clinical Pathology 2002; 55: 435-439.

Noorani HZ, Brown A, Skidmore B et al. Liquid-based cytology and human papillomavirus testing in cervical cancer screening. Canadian Coordinating Office for Health Technology Assessment (CCOHTA) 2003.

Obwegeser KH, Brack S. Does liquid-based technology really improve detection of cervical neoplasia? A prospective, randomized trial comparing the ThinPrep Pap Test with the conventional Pap Test, including follow-up of HSIL cases. Acta Cytologica 2001; 45: 709-714.

Ogilvie GS, Patrick DM, Schulzer M et al. Diagnostic accuracy of self collected vaginal specimens for human papillomavirus compared to clinician collected human papillomavirus specimens: a meta-analysis. Sexually Transmitted Infections 2005.

Olaniyan OB. Validity of colposcopy in the diagnosis of early cervical neoplasia-a review. African Journal of Reproductive Health 2002; 6: 59-69.

Ostor AG. Natural history of cervical intraepithelial neoplasia: a critical review. International Journal of Gynecological and Obstetrics 1993; 12: 186-192.

Paavonen J, Jenkins D, Bosch FX et al. for the HPV PATRICIA study group. Efficacy of a prophylactic adjuvanted bivalent L1 virus-like-particle vaccine against infection with human papillomavirus types 16 and 18 in young women: an interim analysis of a phase III double-blind, randomised controlled trial. The Lancet 2007; 369: 2161-2170.

PAB (Professional Advisory Board) of the Australia and New Zealand HPV Project. Guidelines for the Management of Genital Human Papillomavirus (HPV) in Australia and New Zealand 5th Edition – 2007. Viral Sexually Transmitted Infection Education Foundation 2007.

Palefsky JM, Holly EA. Chapter 6: Immunosuppression and co-infection with HIV. Journal of the National Cancer Institute Monographs 2003:41–6.

Papanicolaou GN. Atlas of Exfoliative Cytology. Harvard University Press, 1954.

Papanicolaou GN.; Trout, HE. Diagnosis of uterine cancer by vaginal smears. New York: The Commonwealth Fund. 1943.

Parboosingh J. Screening for cervical cancer: Canadian programmatic guidelines. Canadian family physician 1999; 45: 383–93.

Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA: a Cancer Journal for clinicians 2002; 55(2):74-108.

Peters RK, Thomas D, Skultin G. Invasive squamous cell carcinoma of the cervix after recent negative cytology test results: A distinct subgroup. American Journal of Obstetrics & Gynecology 1988; 158: 926.

Peterson O. Spontaneous course of cervical precancerous conditions. American Journal of Obstetrics & Gynecology 1956; 7: 1063-1071.

Petignat P, Faltin DL, Bruchim I et al. Are self-collected samples comparable to physician-collected cervical specimens for human papillomavirus DNA testing? A systematic review and meta-analysis. Gynecologic Oncology 2007; 105: 530-535.

Petry KU, Menton S, Menton M et al. Inclusion of HPV DNA testing in routine cervical cancer screening for women above 29 years in Germany: results for 8466 patients. British Journal of Cancer 2003; 88: 1570-1577.

Petterson F, Björkholm E, Näslund I. Evaluation of screening for cervical cancer in Sweden: Trends in incidence and mortality 1958–1980. International Journal of Epidemiology 1985; 14: 521-527.

Pinto AV, Crum CP. Natural history of cervical neoplasia: defining progression and its consequences. Clinics in Obstetrics and Gynecology 2000; 43: 352-362.

Prorok PC, Chamberlain J, Day NE et al. UICC workshop on the evaluation of screening programmes for cancer. International Journal of Cancer 1984; 34: 1–4.

Quinn M, Babb P, Jones J et al. Effect of screening on incidence of and mortality from cancer of cervix in England: evaluation based on routinely collected statistics. British Medical Journal 1999; 318: 904-908.

Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. New England Journal of Medicine 1978; 299: 926-930.

Ratnam S, Franco EL, Ferenczy A. Human papillomavirus testing for primary screening of cervical cancer precursors. Cancer Epidemiology Biomarkers and Prevention 2000; 9: 945-951.

Remmink AJ, Walboomers JMM, Helmerhorst TJM et al. The presence of persistent high-risk HPV genotypes in dysplastic cervical lesions is associated with progressive disease: Natural history up to 36 months. International Journal of Cancer 1995; 61: 306-311..

Richart RM. Natural history of cervical intraepithelial neoplasia. Clinical Obstetrics and Gynecology 1968; 5: 748-784.

Ries LAG, Melbert D, Krapcho M, et al (eds). SEER Cancer Statistics Review, 1975-2004, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_ 2004/, based on November 2006 SEER data submission, posted to the SEER web site, 2007.

Ronco G, Cuzick J, Pierrotti P et al. Accuracy of liquid based versus conventional cytology: Overall results of new technologies for cervical cancer screening: Randomised controlled trial. British Medical Journal 2007b; 335: 28-34.

Ronco G, Cuzick J, Segnan N et al. HPV triage for low grade (L-SIL) cytology is appropriate for women over 35 in mass cervical cancer screening using liquid based cytology. European Journal of Cancer 2007a; 43: 476-480.

Ronco G, Giorgi-Rossi P, Carozzi F et al. Human papillomavirus testing and liquid-based cytology in primary screening of women younger than 35 years: results at recruitment for a randomised controlled trial. Lancet Oncology 2006; 7: 547-555.

Ronco G, Pilutti S, Patriarca S et al. Impact of the introduction of organized screening for cervical cancer in Turin, Italy: cancer incidence by screening history 1992–98. British Journal of Cancer 2005; 93: 376–378.

Rose J, Beaulac J, Howlett R et al. Cervical Cancer Prevention and Control Network. Cervical cancer in Canada: a response to the release of the CCS/NCIC Cancer Statistics 2006. Journal of Obstetrics and Gynecology Canada 2006; 28: 678-679.

Sadler L, Priest P, Peters J et al. The New Zealand Cervical Cancer Audit, Whakamaatau Mate Pukupuku Taiawa o Aotearoa. Wellington: Ministry of Health, 2004a.

Sadler L, Saftlas A, Wang W et al. Treatment of cervical intraepithelial neoplasia and risk of preterm delivery. Journal of the American Medical Association 2004b; 291: 2100–2106.

Saint M, Gidengorin G, Sawaya GF. Current cervical neoplasia screening practices of obstetrician/gynecologists in the US. American Journal of Obstetrics and Gynecology 2005; 192: 414-421.

Salmeron J, Lazcano-Ponce EC, Lorincz A et al. Comparison of HPV-based assays with Papanicolaou smears for cervical cancer screening in Morelos State, Mexico. Cancer Causes & Control 2003; 14: 505-512.

Sangwa-Lugoma G, Mahmud S, Nasr SH et al. Visual Inspection as a cervical cancer screening method in a primary health care setting in Africa. International Journal of Cancer 2006; 119: 1389-1395.

Sankaranarayanan R, Chatterji R, Shastri S et al. Accuracy of human papillomavirus testing in the primary screening of cervical neoplasia: results from a multicentre study in India. International Journal of Cancer 2004; 112: 341–347.

Sankaranarayanan R, Esmy PO, Rajkumar R et al. Effect of visual screening on cervical cancer incidence and mortality in Tamil Nadu, India: a cluster-randomised trial. Lancet. 2007; 370: 398-406.

Sankaranarayanan R, Gaffikin L, Jacob M et al. A critical assessment of screening methods for cervical neoplasia. International Journal of Gynaecology and Obstetrics. 2005; 89 Suppl 2:S4-S12.

Sankaranarayanan R, Wesley R, Thara S et al. Test characteristics of visual inspection with 4% acetic acid (via) and Lugol's iodine (VILI) in cervical cancer screening in Kerala, India. International Journal of Cancer 2003; 106: 404-408.

Saslow D, Castle P, Cox JT et al. American Cancer Society guideline for human papillomavirus (HPV) vaccine use to prevent cervical cancer and its precursors. CA: a Cancer Journal for Clinicians 2007; 57: 7-28.

Sawaya GF, Smith-McCune K. HPV vaccination — More answers, more questions. New England Journal of Medicine 2007; 356: 1991-1993.

Schlecht NF, Kulaga S, Robitaille J et al. Persistent human papillomavirus infection as a predictor of cervical intraepithelial neoplasia. Journal of the American Medical Association 2001; 286: 3106-3114.

Schiffman M. Integration of human papillomavirus vaccination, cytology, and human papillomavirus testing. Cancer Cytopathology 2007; 111: 145-153.

Schiffman M, Bauer HM, Hoover RN, et al. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. Journal of the National Cancer Institute 1993; 85: 958-964.

Schiffman M, Herrero R, Hildesheim A et al. HPV DNA testing in cervical cancer screening - Results from women in a high-risk province of Costa Rica. Journal of the American Medical Association 2000; 283: 87-93.

Schiffman M, Kruger, Kjaer S. Chapter 2: Natural history of anogenital infection and neoplasia. Journal National Cancer Institute Monographs 2003; 31: 14-19.

Schneider A, Hoyer H, Lotz B et al. Screening for high-grade cervical intra-epithelial neoplasia and cancer by testing for high-risk HPV, routine cytology or colposcopy. International Journal of Cancer 2000; 89: 529-534.

Sellors J, Cosby R, Trim K et al. Recruiting family physicians and patients for a clinical trial: lessons learned. Journal of Family Practice 2002; 19: 99–104.

Selvin E, Brett KM. Breast and cervical cancer screening: sociodemographic predictors among White, Black, and Hispanic women. American Journal of Public Health 2003; 93: 618-623.

Shapiro SH, Glass KC. Why Sackett's analysis of randomized controlled trials fails, but needn't. Canadian Medical Association Journal 2000; 163: 834–835.

Shaw PAV. Lessons from the audit of invasive cervical cancer. Cytopathology 2002; 13: 1–3.

Shepherd J, Weston R, Peersman G et al. Interventions for encouraging sexual lifestyles and behaviours intended to prevent cervical cancer. Cochrane Gynaecological Cancer Group Cochrane Database of Systematic Reviews 2005; 2.

Sherman ME, Wang SS, Wheeler CM et al. Determinants of human papillomavirus load among women with histological cervical intraepithelial neoplasia 3: dominant impact of surrounding low-grade lesions. Cancer Epidemiology Biomarkers and Prevention 2003; 12: 1038-1044.

Sideri M, Spolti N, Spinaci L et al. Interobserver variability of colposcopic interpretations and consistency with final histologic results. Journal of Lower Genital Tract Diseases 2004; 8: 212–216.

Silins I, Ryd W, Tornberg S, Strand A et al. Duration of the decreased risk for development of high grade CIN after negative cytology and negative HPV DNA test: a population based cohort study. Oral Presentation C-08, 22nd International papillomavirus Conference, Vancouver, Canada, April 30-May 6, 2005.

Smith RA, Cokkinides V, Eyre HJ. American Cancer Society guidelines for the early detection of cancer, 2006. CA: a Cancer Journal for Clinicians. 2006; 56: 11-25.

Snijders PJ, Hogewoning CJ, Hesselink AT et al. Determination of viral load thresholds in cervical scrapings to rule out CIN 3 in HPV 16, 18, 31 and 33-positive women with normal cytology. International Journal of Cancer 2006; 119: 1102-1107.

Snijders PJ, Van den Brule AJ, Meijer CJ. The clinical relevance of human papillomavirus testing: relationship between analytical and clinical sensitivity. Journal of Pathology 2003; 201: 1-6.

Solomon D, Davey D, Kurman R et al. The 2001 Bethesda System: Terminology for reporting results of cervical cytology. Journal of the American Medical Association 2002; 287: 2114-2119.

Solomon D, Schiffman M, Tarone R. Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: baseline results from a randomized trial. Journal of the National Cancer Institute 2001; 93: 293-299.

Spence AR, Goggin P, Franco EL. Process of care failures in invasive cervical cancer: Systematic review and meta-analysis. Preventive Medicine 2007. In press.

Stoler MH, Schiffman M. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. Journal of the American Medical Association 2001; 285:1500-1505.

Stuart GCE, McGregor E, Duggan MA et al. Review of the screening history of Alberta women with invasive cervical cancer. Canadian Medical Association Journal 1997; 157: 513–519.

Sun CA, Lai HC, Chang CC et al. The significance of human papillomavirus viral load in prediction of histologic severity and size of squamous intraepithelial lesions of uterine cervix. Gynecologic Oncology 2001; 83: 95-99.

Syrjänen K, Hakama M, Saarikoski S et al. Prevalence, incidence, and estimated life-time risk of cervical human papillomavirus infections in a nonselected Finnish female population. Sexually Transmitted Diseases 1990; 17: 15-19.

Tacken MA, Braspenning JCC, Hermens RPMG et al. Uptake of cervical cancer screening in The Netherlands is mainly influenced by women's beliefs about the screening and by the inviting organization. European Journal of Public Health 2007; 17: 178-185.

Taube A. Sensitivity, specificity and predictive values: a graphical approach. Statistics in Medicine 1986; 5: 585-591.

Trottier H, Franco EL. The epidemiology of genital human papillomavirus infection. Vaccine 2006; 24: S4-S15.

Vaccarella S, Franceschi S, Herrero R et al. Sexual behavior, condom use, and human papillomavirus: Pooled analysis of the IARC human papillomavirus prevalence surveys. Cancer Epidemiology Biomarkers & Prevention 2006; 15: 326-333.

Van Ballegooijen M, van den Akker-van Marle E, Patnick J et al. Overview of important cervical cancer screening process values in European Union (EU) countries, and tentative predictions of the corresponding effectiveness and cost-effectiveness. European Journal Cancer 2000; 36: 2177-2188.

Villa LL, Costa RL, Petta CA et al. Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. Lancet Oncology 2005; 6: 271-278.

Villa LL, Costa RL, Petta CA et al. High sustained efficacy of a prophylactic quadrivalent human papillomavirus types 6/11/16/18 L1 virus-like particle vaccine through 5 years of follow-up. British Journal of Cancer. 2006 Dec 4; 95: 1459-1466.

Walboomers JMM, Jacobs MV, Manos MM et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. Journal of Pathology 1999; 189: 12-19.

Waller J, McCaffery K, Forrest S et al. Awareness of human papillomavirus among women attending a well woman clinic. Sexually Transmitted Infections 2003; 79: 320-322.

Waller J, McCaffery K, Kitchener H et al. Women's experiences of participation in repeated testing for human papillomavirus (HPV) in the context of cervical cancer screening: a qualitative study. Psychology 2006; 16: 196-204.

Waller J, McCaffery K, Nazroo J et al. Making sense of information about HPV in cervical screening: a qualitative study. British Journal of Cancer 2005; 92: 265-270.

Wallin KL, Wiklund F, Angstrom T et al. Type-specific persistence of human papillomavirus DNA before the development of invasive cervical cancer. New England Journal of Medicine 1999; 341: 1633-1638.

Winer RL, Hughes JP, Feng Q, et al. Condom use and the risk of genital human papillomavirus infection in young women. New England Journal of Medicine 2006; 354: 2645-2654.

WMA. World Medical Association. Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects 2004. Accessed July 2007 at http://www.wma.net/e/policy/b3.htm.

von Knebel-Doeberitz M, SyrjänenKJ. Molecular markers: How to apply in practice. Gynecologic Oncology 2006; 103:18-20.

Woolf SH, Battista RN, Anderson GM, et al. and the Canadian Task Force on the Periodic Health Examination. Assessing the clinical effectiveness of preventive maneuvres: analytic principles and systematic methods in reviewing evidence and developing clinical practice recommendations. Journal of Clinical Epidemiology 1990; 43: 891-905.

Wright TC. Chapter 10: Cervical cancer screening using visualization techniques. Journal National Cancer Institute Monographs 2003b; 31: 66-71.

Wright TC, Cox JT, Massad LS et al. Guidelines for the management of women with cervical cytological abnormalities - ASCCP-sponsored Consensus Conference. Journal of the American Medical Association 2002; 287: 2020-2029.

Wright TC, Cox JT, Massad LS et al. Society for colposcopy and cervical pathology. 2001 Concensus Guidelines for the management of women with cervical intraepithelial neoplasia. American Journal of Obstetric and Gynecology 2003a; 189: 295-304.

Wright TC, Schiffman M. Adding a test for human papillomavirus DNA to cervical-cancer screening New England Journal of Medicine 2003c; 348: 489-490.

Wright TC, Schiffman M, Solomon D et al. Interim guidance for the use of human papillomavirus DNA testing as an adjunct to cervical cytology for screening. Obstetrics and Gynecology 2004; 103: 304-309.

Youden WJ. Index for rating diagnostic tests. Cancer 1950; 3: 32-35.

Zhou XH. Comparing accuracies of two screening tests in a two phase study for dementia. Applied Statistics 1998; 47: 135-147.

Zhou XH, McClish DK, Obuchowski NA. Statistical Methods in Diagnostic Medicine. John Wiley and Sons Inc., New York, USA, 2002, 464 pages.

Zielinski GD, Snijders PJF, Rozendaal L et al. High-risk HPV testing in women with borderline and mild dyskaryosis: Long-term follow-up data and clinical relevance. Journal of Pathology 2001; 195: 300-306.

Zur Hausen H. Viruses in human cancers. Science 1991; 254: 1167-1173.

APPENDIX A

Permission to use Figure from Baldwin P. Laskey R. Coleman N. Translational approaches to improving cervical screening. Nature Reviews. Cancer. 3: 217-226, 2003. as Figure 4 in thesis.

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Jul 26, 2007

This is a License Agreement between Marie-Helene Mayrand ("You") and Nature Publishing Group ("Nature Publishing Group"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

License Number	1756560588959
License date	Jul 26, 2007
Licensed content publisher	Nature Publishing Group
Licensed content publication	Nature Reviews Cancer
Licensed content title	Opinion: Translational approaches to improving cervical screening
Licensed content author	Peter Baldwin, Ronald Laskey, Nicholas Coleman
Volume number	
Issue number	
Pages	
Year of publication	2003
Portion used	Figures
Number of figures	1
Requestor type	Student
Type of Use	Thesis / Dissertation
PO Number	
Total	0.00
Terms and Conditions	

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use, subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version. Where print permission has been granted for a fee, separate

permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run). NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.
5. The credit line should read:

   Reprinted by permission from: Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   For AOP papers, the credit line should read:

   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

   Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedia.com/communications.com for more information. Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:

   Translated by permission from Macmillan Publishers Ltd: JOURNAL NAME] *(reference citation)*, copyright (year of publication).

We are confident that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

v1.1

EMAIL EXCHANGE GRANTING PERMISSION TO USE PAP PICTURE IN FIGURE 5.

Yes you have my permission to use the pap smear image in your thesis.

Take Care and Think Microbiologically!
Neal R. Chamberlain, Ph.D.
Associate Professor
Department of Microbiology/Immunology
A.T. Still University of Health Sciences/KCOM
Kirksville, Missouri
http://www.kcom.edu/faculty/chamberlain

>>> "M.-H. Mayrand" <mhmayrand@videotron.ca> 8/7/2007 10:02 AM >>>

Hello!

I am a PhD student at McGill University (Montreal, Canada) and I am ready to submit my thesis. I would like to include a picture of a Pap test. I found one at the following link, with your name next to the copytright:

http://www.kcom.edu/faculty/chamberlain/Website/lectures/lecture/image/dyspl.jpg

Would you let me include it in my thesis?

Thank you

Marie-Helene Mayrand

--- Scanned by M+ Guardian Messaging Firewall ---

WEBSITE TEXT GIVING PERMISSION TO USE PAP SAMPLING PICTURE IN FIGURE 5.

The text can be found at : http://lib-sh.lsuhsc.edu/fammed/default.html#Atlas

Introduction to colposcopy
Sponsored and developed by
the Department of Family Medicine and Comprehensive Care,
Louisiana State University Health Sciences Center
Shreveport, Louisiana

The materials that follow are meant to be used by physicians, medical students, and health educators for medical training, medical practice, and patient education. Please use these materials freely

The picture used can be found at http://lib-sh.lsuhsc.edu/fammed/atlases/ecc.jpg

EMAIL EXCHANGE GRANTING PERMISSION TO USE HPV TECHNOLOGY PICTURES IN FIGURE 6.

Dr. Mayrand –

On behalf of Digene, I give you permission to use the images shown from the Digene.com website for your thesis. Specifically, the three images to be used are from the illustration of an overview of Hybrid Capture technology, as shown below.

This is for one-time use only and all other uses are prohibited without prior consent.

Please also provide credit for the images as follows: © *Digene Corporation – used by permission.*

Thanks,

Shelley Ducker

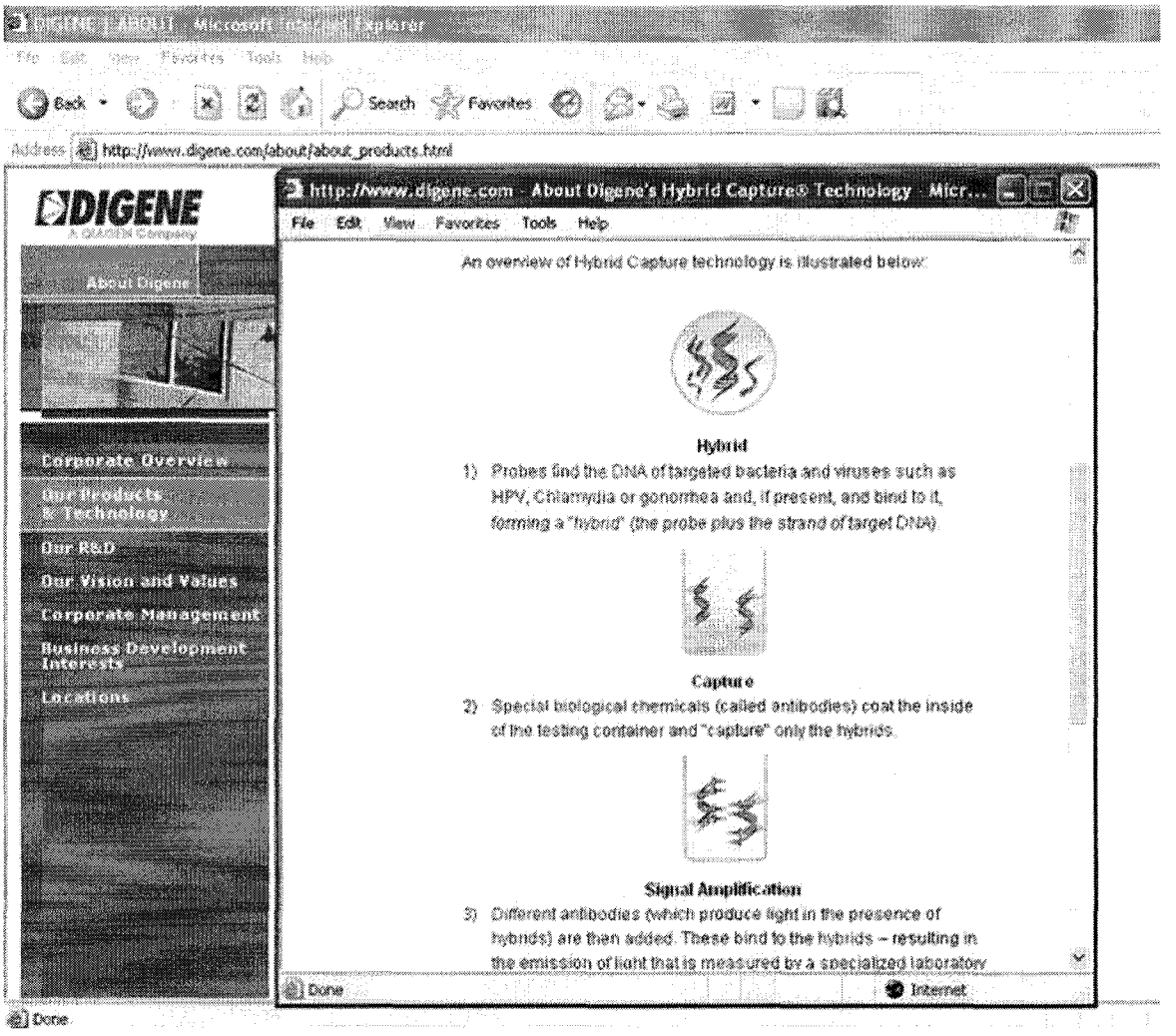
Associate Director, Corporate Communications

Digene Corp.

w: 301.944.7610

c: 202.255.0561

e: shelley.ducker@digene.com



APPENDIX B

CCaST Study

Patient Informed Consent Form

Research Project: CCaST Study (Cervical cancer screening trial): Efficacy trial of HPV versus Pap testing for screening for cervical cancer precursors.

Principal investigator: Dr. Eduardo Franco

Should you decide to participate in this study, you will be given a copy of this consent form; it provides you with a detailed description of the study, describing all the procedures that will be followed. If you have any questions concerning what is explained here, don't hesitate in asking us. Please, take all the time you need to read this form.

1. Purpose of the study

You have been asked to take part in a study examining the value of Pap smear compared to human papillomavirus (or HPV, for short) testing in the screening of cervical cancer. We will enrol a total of 12,000 women in this study.

The cervix is the opening of the uterus. Regular Pap tests can prevent a substantial portion of cervical cancers. The Pap test, or cervical cytology, is the test that your doctor collects during a routine gynecologic examination. The sample collected is sent to a laboratory and examined under a microscope in order to detect precancerous cells. These cells can then be treated and cancer can be prevented. We now know that precancer and initial cancer changes on the cervix are caused by certain types of HPV. We hope that by testing for cancer-causing HPV we will be able to identify even more women who have precancer changes on their cervix.

2. Study design

If you consent to take part in this study, you will have a 50:50 chance, as in the flip of a coin to be assigned to either:
(1) Pap smear followed by HPV testing or
(2) HPV testing followed by Pap smear.

Both tests will be collected during the same gynecologic examination, which you would have had anyway today. The collection of the second test takes around 30 seconds. You will have a Pap test done in the usual way. The specimen collected for HPV testing will be sent to the laboratory and will be tested for the most common types of HPV that cause cervical cancer. As part of laboratory quality control, we will send some specimens to be analysed in other laboratories in order to confirm our results. We also ask your permission to keep this sample for future studies about conditions that affect cervical infection and cervical disease using more refined technologies not included in this study.

Colposcopy examination

When a woman has an abnormal Pap smear or HPV test, we cannot be sure if she has precancer or not. This is why she is usually referred for a more precise test: a colposcopy. This is an examination of the cervix with a magnifying lens. The physician performing the colposcopy can look for small areas of precancerous changes on the cervix. He can take biopsies (small pieces of the cervix) that the pathologist will examine. Colposcopy is considered the most sensitive test to detect precancerous and cancerous lesions of the cervix.

Although the Pap smear and the HPV test are good tests, they are not perfect. Sometimes they will miss precancer changes, and sometimes they can mistake normal changes for precancer changes. To evaluate how frequent these mistakes are, we will ask certain women with normal Pap smear and HPV test results to go to colposcopy. This examination will be done in a colposcopy clinic in the same area as the physician you are visiting today. This examination will take approximately 20 minutes. If you have normal results on both tests you have a 1 on 5 chance of being asked to undergo colposcopy.

Follow-up

If the result of both tests are normal and you were not selected for a colposcopic examination (as described above, you will be invited to return in one year to repeat both tests, the Pap cytology and the HPV test.

If all your tests were normal and you were selected to undergo a colposcopic examination which yields normal results, you will be invited to return in one year to repeat the Pap cytology and the HPV test.

3. Benefits

If you agree to participate in this study, you will have access to the best screening methods available for cervical cancer. This may not benefit you directly, if you do not have precancerous lesions of that cervix. However, your participation will contribute to a better understanding of cervical cancer screening, which may benefit you at a later time in your life, and that will certainly benefit other women.

4. Risks

The Pap test will be collected as usual. There are no risks involved when a Pap smear is taken. Even in the very unlikely event of bleeding on the site of collection of the smear, this is never significant and does not require additional treatment.

Most HPVs that cause cancer of the cervix are sexually transmitted. These sexually transmitted viruses are very frequent. So much so, that up to 50% of women will have this type of infection at one point of her life or another. Fortunately, more than 99% of women who have this virus will never get cervical cancer. Most infections go away by themselves and do not cause precancer or cancer. However, some will. This is why in this study, women who are positive for HPV will be asked to undergo colposcopic examination. Those women with a normal examination should be reassured that their infection is not causing precancer or cancer changes. However, it is possible that some women may worry that they are at risk for future cancer, and some women may be upset that they have an infection that could have been sexually transmitted. Counselling will be available to all women who request it or who appear to be under stress. It should be stressed that an HPV infection, once acquired can persist for a very long time. Thus, a positive test for an HPV infection is in no way a sign of a recent change of sexual partner.

When a biopsy is taken at colposcopy there is a small (less than 1 in 1000) risk of bleeding or infection. These usually go away without treatment. Nonetheless, you will be provided with telephone numbers of study personnel that can be contacted 24 hours a day should you need to consult a physician.

5. Confidentiality

The results from the analyses of your cervical samples, as well as the responses you gave to the questionnaire will be treated in strict confidentiality. All the data from this study will be analyzed in aggregate statistical form only, again with no names linked to any data.

The actual samples from your cervical cells will not be made available to investigators that are not involved with this study, nor will they be sold for commercial use. They will only be used for the purposes outlined in this consent form. They will be securely stored for as long as they are needed for the verification of laboratory results, testing with auxiliary methods, and for research audit purposes. Your name will not be linked to any specimen.

6. Withdrawal from the study
Your participation in this study is completely voluntary. You are free to withdraw from the study at any time. Your decision to withdraw will have no effect on your current or future medical care.

7. Ethical approval
The Institutional Review Board of the McGill University has approved this research project.

Patient Initials: _____

Research Project: CCaST: Cervical cancer screening trial: Efficacy trial of HPV versus Pap testing for screening for cervical cancer precursors

Principal investigator: Dr. Eduardo Franco

CONSENT

My signature on this form indicates that I have understood to my satisfaction the information regarding my participation in this research project and agree to participate as a study subject. In no way does this waive my legal rights nor release the investigators, nor involved institutions from their legal and professional responsibilities. I am free to withdraw from this study at any time. My continued participation should be as informed as my initial consent, so I am free to ask for clarification or new information throughout my participation. I understand that if I have any questions concerning matters related to this research, I may call Dr. Eliane Duarte-Franco, Research Coordinator at 514-398-5543.

Name of participant	Signature of participant	Date

Name of witness	Signature of witness	Date

APPENDIX C

CCaST Study

ENROLLMENT FORM

Form No: _____ Today's Date: _____/_____/_____
Day Month Year

This clinic is taking part in a study of screening for cervical cancer. Please answer the questions below, even if you are not interested in the study.

1.	Do you have a problem understanding English?	❑ No	❑ Yes
2.	To the best of your knowledge, have you had a surgery to remove your cervix or uterus (a hysterectomy)?	❑ No	❑ Yes
3.	To the best of your knowledge, are you pregnant?	❑ No	❑ Yes
4.	During the past 2 years, have you had a colposcopic examination?	❑ No	❑ Yes
5.	Are you under 30 years of age?	❑ No	❑ Yes
6.	Are you over 69 years of age?	❑ No	❑ Yes
7.	Did you have a Pap test less than 1 year ago?	❑ No	❑ Yes
8.	Do you participate in the CCaST study?	❑ No	❑ Yes

If you answered **"No"** to all 8 questions above you are eligible to participate in the CCaST study. Would you like to participate in this study and help us answer important questions on cervical cancer screening?

❑ Yes ⇩

❑ No ⇨ THANK YOU FOR ANSWERING THESE QUESTIONS

PLEASE, PROVIDE THE FOLLOWING INFORMATION:

Your name: _____

Your address: _____

_____

Phone # (home): _____

Phone # (work/other): _____

Phone # (cellular): _____

Phone # (relative/friend): _____

Medical Chart #:

Doctor's name:

ADDRESSOGRAPH

AFTER COMPLETING THIS FORM, PLEASE RETURN IT TO THE PERSON WHO GAVE IT TO YOU, TO YOUR NURSE OR DOCTOR.
For detailed information on the study or for general information on screening for cervical cancer, you will find pamphlets on display in this waiting room; please, feel free to take a copy, or speak with your doctor or nurse.
THANK YOU FOR YOUR PRECIOUS COLLABORATION !

For the use of the CCAST study assistant only:
WRITE OR PASTE THE LABEL WITH THE PATIENT'S STUDY NUMBER HERE. **CCaST Study No.** _____

CCaST Study

PATIENT QUESTIONNAIRE - ENROLMENT

INSTRUCTIONS FOR THE QUESTIONNAIRE

This questionnaire contains questions on general information, your medical history and about the steps you took in order to attend this visit.

Most questions require that you simply check a box ❑ with an "X" to indicate your choice. Others require a specific answer, such as age, date, or another number. Depending on your answer for some questions, you will be told to skip ahead to a different part of the questionnaire. This is to save you time, so that you won't have to go over questions that do not apply to you.

There is no right or wrong answer to any question. Many questions require that you think back over your adult years, particularly over the past year, to recall specific information. Please take the time to reflect. If you can't possibly remember the information skip the question, but we would like to encourage you to try to answer all questions. A good guess is always better than no information at all. If you would like to tell us more about any specific item, please use the space available at the end of the questionnaire.

WE VERY MUCH APPRECIATE YOUR COLLABORATION IN THIS STUDY!

STUDY NO: _____ TODAY'S DATE: _____/_____/_____
DAY MONTH YEAR

YOUR INITIALS: _____

GENERAL INFORMATION

1. How old are you? Age: _____ years Date of birth: _____/_____/_____ DAY MONTH YEAR

2. Your current marital status (choose the ONE that best describes your actual situation):
   - ❑ Married
   - ❑ Single
   - ❑ Unmarried, but living with a partner
   - ❑ Divorced/separated
   - ❑ Widowed

3. Your current marital status (choose the ONE that best describes your actual situation):
   - ❑ Married
   - ❑ Single
   - ❑ Unmarried, but living with a partner
   - ❑ Divorced/separated
   - ❑ Widowed

4. What ethnic background do you consider yourself to belong?
   - ❑ French/French Canadian
   - ❑ Anglo/English Canadian
   - ❑ Black Canadian
   - ❑ Chinese
   - ❑ Asian/Oriental (not Chinese)
   - ❑ Native Indian
   - ❑ Hispanic
   - ❑ Portuguese
   - ❑ Greek
   - ❑ Jewish
   - ❑ Italian
   - ❑ Arab
   - ❑ Other:_____

5. Your current occupation is: _____

6. What is the highest level of schooling you attained?
   - ❑ Elementary school
   - ❑ High School (incomplete)
   - ❑ High School (complete)
   - ❑ College
   - ❑ University (undergraduate)
   - ❑ University (Baccalaureate or higher)

7. Have you ever smoked cigarettes regularly, (one cigarette or more every day for a year or more)?
   - ❑ Yes
   - ❑ No
     If **No**, go to question **9**

8. Do you still smoke?
   - ❑ Yes
   - ❑ No
     If **No**, at what age did you stop? Age: _____ years

MEDICAL, GYNECOLOGIC & OBSTETRIC HISTORY

9. When did you have your last menstrual period? _____ / _____ MONTH YEAR

10. Have you undergone menopause yet?

❑ Yes ❑ No ❑ I don't know

If **No** OR if **you don't know**, go to question **12**

11. Are you currently taking hormonal supplements that were prescribed by a doctor (in the form of pills, patch, cream or gel)?

❑ Yes ❑ No ❑ I don't know

12. How old were you the first time you had sexual intercourse?

Age: _____years ❑ I have never had sexual intercourse

If **Never**, go to question **19**

13. What is the number of male partners with whom you have ever had sexual intercourse?

Number *(approximately)* _____

14. During the last year only, what is the number of male partners with whom you have had sexual intercourse?

_____ Number ❑ None in the past year

→ How many of those partners were new?_____ Number

15. What is the method of contraception you are using now?

❑ None
❑ Depo-Provera or Norplant
❑ Condom
❑ Other (please specify): _____
❑ Birth control pill
❑ Surgical: vasectomy, tubal ligation
❑ IUD (intra-uterine device)

16. To the best of your knowledge, are you currently pregnant?

❑ Yes ❑ No ❑ I don't know

17. Have you ever been pregnant before?

❑ Yes ❑ No

If **No**, go to question **19**

If **Yes**,

How many times (including premature births, full-term births, stillbirths, ectopic pregnancies, miscarriages and abortions)? _____ times

18. How old were you when you became pregnant for the first time? Age: _____ years

19. Have you ever had a Pap smear (cervical cytology) taken before?

❑ Yes ❑ No, this is my first Pap smear ❑ I don't know

If this is your **first Pap smear** OR if **you don't know**, go to question **22**

20. When did you have your last Pap smear taken? _____/_____

MONTH YEAR

21. Have you ever been told that your Pap smear was abnormal?

❑ Yes ❑ No ❑ I don't know

YOUR PERSONAL FINANCIAL IMPLICATIONS CONCERNING THIS MEDICAL VISIT

22. Did you miss work to attend this medical appointment?

❑ Yes ❑ No

If **No**, go to question **25**

23. How many hours of work did you miss today to attend your appointment? _____ hours

24. Did the number work hours you missed affect your pay or was the time granted by your employer?

❑ it affected my pay ❑ It was time granted by my employer

25. Estimate the total time you had to devote to this appointment, including transportation, waiting time, meeting with the doctor.

_____ hours and _____ minutes

26. What means of transportation did you use to come to this appointment?

❑ Private: (car) Estimate the round trip distance in kilometers: _____km

❑ Public transportation (bus, metro, train, taxi)

27. Quantify your expenses ensuing from the present appointment.
(Please, enter "**0**" if you had no expenses):

❑ Parking: $ _____

❑ Public transportation: Taxi: $ _____

Bus/Metro/Train: $ _____

❑ Babysitter: $ _____

❑ Other: _____ $ _____

(Example: meals away from home)

28. What is your usual yearly household income?

❑ < $10,000 (OR < $850/month)

❑ $10,000 to $ 20,000 (OR $850 to <$1,500/month)

❑ $20,000 to $ 40,000 (OR $1,500 to <$3,500/month)

❑ $40,000 to $ 60,000 (OR $3,500 to <$5,000/month)

❑ $60,000 to $ 80,000 (OR $5,000 to <$6,500/month)

❑ > $ 80,000 (OR >= $6,500/month)

END OF QUESTIONNAIRE **Thank you very much!**