# 1. INTRODUCTION

In most of developed countries, the incidence and mortality of cervical cancer has significantly decreased after introduction of routine Pap smear screening. Australia has the second lowest incidence of cervical cancer and the lowest mortality from cervical cancer in the world (Australian Institute of Health and Welfare (AIHW) 2006). It has been a long debate on the best possible use of resources for cervical cancer screening. Suggestions have been made for extending the screening interval or commencing the screening at later ages achieving substantial savings which could be spent on targeting non-participants (Carter, Stone et al. 2000). An international comparison study on cervical cancer screening also suggested that Australia, could reduce the cost of cervical cancer with no effect on health outcomes, suggesting there are efficiencies to be gained (van den Akker-van Marle, van Ballegooijen et al. 2002). However, the possibility of reducing the current mix of services by strategies such as reducing the screening interval have been met by significant opposition from various key stakeholders (National Cancer Control Initiative 2001). In the past 10 years, great technological advances have been made including, the Human Papillomavirus (HPV) DNA test in the detection of cervical cancer and the HPV vaccine in the prevention of cervical diseases. The understanding of the natural history of cervical cancer and the new technologies has brought a promising future for further cervical cancer control.

The present study has been undertaken within the context of the ACE-Prevention study. ACE-Prevention aims to evaluate the cost-effectiveness of 100 preventative interventions for non-communicable disease and compare these to 50 benchmark interventions in order to inform Australian health policy about efficient preventative interventions. The primary comparator of all ACE-Prevention studies is current practice though the null scenario is also considered as per the WHO-CHOICE generalised cost-effectiveness methodology (Tan-Torres Edejer, Baltussen et al. 2003). The perspective of ACE-Prevention is largely the health sector with costs divided into costs to government and costs to patients and the reference year is 2003. The key health outcome measure of all studies is the disability-adjusted life year (DALY). All ACE-studies are guided by a comprehensive and detailed economic protocol which is publicly available and ensures all studies are methodologically congruent (hence avoiding confounding). The present study aims to evaluate the incremental cost-effectiveness of five screening strategies (detailed below) for cervical cancer compared to the null scenario and current practice.

# 2. METHODOLOGY

In order to evaluate the incremental cost-effectiveness ratios (ICERs) of the five screening strategies, an epidemiological model was developed which details current practice and calculates DALYs gained for each of the six screening strategies. The costs of screening and treatment were determined using detailed pathway analysis.

## 2.1 Current Practice

Current practice for cervical cancer screening is a 2-yearly screening by conventional Pap test delivered through the National Cervical Screening Program (NCSP) which has been implemented since 1991. All women aged 18 to 69 who have ever been sexually active and have no symptoms or history suggestive of cervical pathology should receive routine Pap screening biannually. Women aged over 70 who have never had a Pap smear, or who request a Pap smear, should be screened.

The participation rate of the NCSP has been monitored and reported using the State and Territory Cervical Cytology Registry data. The total percentage of women screened in a 24-month period was 60.7% (95%CI: 60.6%-60.7%) (Australian Institute of Health and Welfare (AIHW) 2006).

## 2.2 Definition of intervention/s

Three screening tools are considered in the current study:

- Conventional Pap smear
- Human Papillomavirus (HPV) DNA test (Hybrid Capture II); and
- Combination of both test (positive on either test will be referred for further investigation)

It is assumed that the NCSP remains the service provider financed by Commonwealth and State governments and the operation of the program is in steady state.

Five screening strategies, variants of the above screening tools, as outlined in Table 1, are considered. These include one decrement from current practice (strategy 1) and four interventions using the new tool of HPV test. Strategy 2 and 3 are simply to screen current target population with extended interval (3 years) by HPV test and combined test. Strategy 4 and 5 are to further explore the optimal utilization of combined test by means of adopting

HPV test to women age older than 30 (Strategy 4) in addition to Pap screening every 3 year for current target population, and by applying combined tests to all women commencing screening from 25 years old (Strategy 5). HPV vaccination, recently recommended by the Pharmaceutical Benefit Advisory Commission and funded by the Australian Government as part of the National Immunisation Program, to prevent cervical cancer is not included in the present analysis.

Table 1: Interventions of interest in the present analysis

| Intervention | Screening tool | Screening interval | Target population |
|---|---|---|---|
| Current Practice | Conventional Pap smear | 2 year | Women aged 18 to 69 |
| Strategy 1 | Conventional Pap smear | 2 year | Women aged 25-69 |
| Strategy 2 | HPV DNA test | 3 year | Women aged 18 to 69 |
| Strategy 3 | Combined test | 3 year | Women aged 18 to 69 |
| Strategy 4 | Conventional Pap smear | 3 year | Women aged 18 to 30 |
| | Combined test | | Women aged 30 to 69 |
| Strategy 5 | Combined test | 3 year | Women aged 25 to 69 |

## 2.3 Health states/risk factors affected by the intervention

The incidence of invasive cervical cancer, including squamous, adenocarcinoma, adenosquamous cancer, was 7.0 per 100,000 women in 2002 (Australian Institute of Health and Welfare (AIHW) 2006) and the mortality was 1.9 per 100,000 women in 2004. The incidence and mortality rates have continuously declined from 13.6 per 100,000 in 1983 (Australian Institute of Health and Welfare (AIHW) 1998) and 4.7 per 100,000 women in 1984, respectively. The investment in the NCSP has cut mortality by 60% and has halved the number of new cases. The survival from invasive cervical cancer is 64% at 5 years, compared to 95% from micro-invasive cervical cancer (Christine H. Holschneider 2006) and 100% for pre-cancerous lesions (Reich, Pickel et al. 2001). Therefore, early detection of cervical cancer and its pre-cancerous lesion leads to significant health gains.

The association of persistent HPV infection and the development of cervical neoplasia has been well documented (Bosch, Lorincz et al. 2002). HPV DNA was detected in 93% of invasive cervical cancer samples world wide (Bosch, Manos et al. 1995). The most prevalent type found in the tumour is HPV 16, following by HPV 18, 45, and 31. Furthermore, HPV infection has been found in 46% of sexually active young women (Bauer, Ting et al. 1991). Infection at younger ages is transient and the majority of infection resolves spontaneously within 24 months (Ho, Bierman et al. 1998). The HPV infection prevalence statistically decreases in older women, which indicates the infection at older age is persistent placing older women at greater risk of cervical cancer (Melkert, Hopman et al. 1993).

## 2.4 Efficacy/Effectiveness of interventions

The efficacy of cancer screening intervention is determined by:

- the screening tools' performance (sensitivity and specificity)
- the screening interval and target population
- the participation of the population (including target and non-target).

Test with good sensitivity can detect the disease at earlier stages to reduce the mortality and morbidity as well as potentially reduce cost associated with later disease stage. Test with good specificity can avoid unnecessary investigation cost arisen from false positive cases. Meta-analyses have been conducted for conventional Pap smear and HPV test. The present study uses a method by Egger (Egger, Smith et al. 2001) to pool screening test accuracy of conventional Pap smear, HPV test, and combination of both for data required by the model but not reported by meta-analyses in the literature. Details of the method and pooled results are attached in Appendix I. The base case values of sensitivity and specificity are outlined in Table 2. In summary, the HPV test is more sensitive in detecting cervical pre-cancerous lesions but less specific than the conventional Pap smear. The negative predictive value (NPV) of the combined test for women older than 30 is virtually 100% which warrants its usefulness in screening older women with extended interval.

## 2.5 Modelling to health outcomes

### The Cervical Cancer Screen Model

A stochastic simulation model, called the ACE Cervical cancer screen model, was developed for the use in the present analysis. The model uses a continuous algorithm to simulate the natural history of cervical cancer and distinguishes various stages along the cancer growth path. In addition to the non-observed true cancer stages, the model distinguishes clinical stages which are empirically observed and comparable to data reported by the cancer registries. By allowing one member of the cohort into the model at a time, the model can track individual life histories which vary from person to person. The intervention uses the same life history as the null scenario but adds the screening parameters. The model is simulated in a single life table cohort for 1,000,000 women and compares each intervention to the null scenario. In order to conduct the incremental analysis for each intervention compared to current practice, a "common random number"

generator is implemented to ensure the null for each intervention is exactly the same. A detailed explanation of the model can be found in Appendix II.

**Staging of cervical cancer**

The Australian Cervical Cytology Registries use 13 histological stages of cervical cancer and its precancerous lesions. These grading are too numerous for modelling since they greatly add to the complexity of the modelling. Instead, the present study uses a five-group staging, i.e. normal/benign, low-grade squamous intraepithelial lesion (LSIL), high-grade squamous intraepithelial lesion (HSIL), micro-invasive cervical cancer, and invasive cervical cancer. The staging is to minimally distinguish cervical abnormalities by differences in treatment and survival but not to create unnecessary modelling complexity. Details of our staging mapped onto the 13 staging states of Cervical Cytology Registries are outlined in Appendix III.

**Model Inputs**

The parameters which operate the disease process and determine the intervention effectiveness are summarised in Table 2. Individual disease progression is governed by a growth variable in a non-linear distribution. The average time to progress for LSIL and HSIL is presented in Table2 and the time is shorter if taken repression into account. LSIL and HSIL can regress in the model if the time until regression randomly drawn is shorter than the time to progress. The time to regression is age dependent, shorter in young women and increasing with age. Details of progression and regression of the disease can be found in the Appendix II. The screening test sensitivity and Specificity are derived from the pooled results discussed in the Appendix I. The population eligible for screening is adjusted by the hysterectomy rate reported by the National Health Survey 2004-5 (Australian Bureau of Statistics 2006). The screening program participation rates are based on the NCSP monitoring report (Australian Institute of Health and Welfare (AIHW) 2006).

The disability weights for health states of cervical cancer and its precancerous lesions are based on the Australian Burden of Disease (BoD) study (Beggs, Vos et al. 2007), and the disability weight of the pre-cancerous lesion is based on a published study (Kulasingam, Hughes et al. 2002). (Refer to Table 2) The Australian BoD study assigns a 0 disability weight to the recovered state of cervical cancer, as clinicians judge the impact of cervical cancer without signs of recurrence five years after detection to be trivial. To be consistent and comparable with the weight of recovered state, a 0 disability weight is assumed for LSIL, suggestive of a negligible impact on women's health in accordance with the current NHMRC guideline's of observational management for LSIL.

Table 2: Model parameters in base case value

| Model parameters | Value | Source |
|---|---|---|
| Sensitivity of Pap to detect | | Appendix I, sources are cited |
| LSIL | 0.403 | |
| HSIL, | 0.628 | |
| Micro-invasive & invasive cancer | 1.00** | |
| Specificity of Pap test | 0.96 | |
| Sensitivity of combined test to detect | | (Cuzick, Szarewski et al. 2003) (Petry, Menton et al. 2003) (Salmeron, Lazcano-Ponce et al. 2003) |
| LSIL | 0.987* | |
| HSIL, | 0.987 | |
| Micro-invasive & invasive cancer | 1.00*** | |
| Specificity of Pap test | 0.914 | |
| Sensitivity of HPV DNA test to detect | | (Cuzick, Clavel et al. 2006) |
| LSIL, overall | 0.713 | |
| HSIL, overall | 0.961 | |
| age <35 | 0.972 | |
| 35-49 | 0.939 | |
| >50 | 0.975 | |
| Micro-invasive & invasive cancer, overall | 0.961 | |
| age <35 | 0.968 | |
| 35-49 | 0.937 | |
| >50 | 1.00 | |
| Specificity of screening test, overall | 0.916 | |
| age <35 | 0.874 | |
| 35-49 | 0.933 | |
| >50 | 0.945 | |
| Disease progression | | Appendix III |
| Average time from LSIL to HSIL | | |
| age<65 | 7.2 years | |
| >65 | 9.4 years | |
| Average time from LSIL to micro-invasive | | |
| age<65 | 16.2 years | |
| >65 | 20.9 years | |
| Disease regression | | Appendix III |
| From LSIL to normal, mean age 20 | 1.75 years | |
| 40 | 8.4 years | |
| >85 | 28.3 years | |
| from HSIL to LSIL, mean age 20 | 2.2 years | |
| 40 | 11.1 years | |
| >85 | 31.1 years | |
| Disability weight for health adjusted life years | | (Beggs, Vos et al. 2007) (Kulasingam, Hughes et al. 2002) |
| LSIL | 0 | |
| HSIL | 0.05 | |
| Micro-invasive | 0.2 | |
| Invasive | 0.43 | |
| Distant | 0.75 | |
| Disease-free | 0.2 | |
| Recovered | 0 | |
| Terminal | 0.93 | |

| age | Hysterectomy rate | Participation rate | (Australian Bureau of Statistics 2006) |
|---|---|---|---|
| 20-24 | 2.7% | 48.9% | (Australian Institute of Health and Welfare (AIHW) 2006) |
| 25-29 | 2.7% | 59.2% | |
| 30-34 | 2.7% | 63.5% | |
| 35-39 | 6.0% | 64.1% | |
| 40-44 | 6.0% | 64.3% | |
| 45-49 | 4.0% | 65.7% | |
| 50-54 | 4.0% | 63.5% | |
| 55-59 | 1.7% | 66.2% | |
| 60-64 | 1.7% | 56.7% | |
| 65-69 | 1.7% | 48.8% | |
| 70-74 | 1.7% | 18.1% | |
| 75-79 | 1.7% | 7.0% | |
| 80+ | 1.7% | 2.2% | |

* Only histology threshold of CIN2/3+ was reported in the literature. Sensitivity of CIN1 was assumed the same as CIN2/3+.
** Sensitivity of Pap to detect micro-invasive & invasive cervical cancer was pooled from 2 studies' data, Cuzick et al (2003) and Guerra et al (1998).
*** Sensitivity of combined test to detect micro-invasive & invasive cervical cancer was assumed to be 100% (because of Pap 100% sensitivity to detect invasive cancer).

## 2.6 Costs of interventions and offsets

Costs of cervical cancer screening are assessed in two components, (a) screening program costs; and (b) costs associated with abnormal screening results. Patient time and travel cost have not yet been included in the costing, and discounting has been carried out but not presented in the results because the health benefit has not yet been discounted. Cost offsets are measured as the total cost associated with an abnormal screening result (the number of abnormal cases is multiplied by the average cost for each stage of cervical abnormality). Cost offsets are achieved by detecting the cervical abnormalities in less severe early stages with less cost associated with assessment, treatment, and follow-up.

### Screening program costs

The key costs of screening comprise: GP consultation (MBS item 36), pathology cost (e.g. cytology of cervical smear and HPV test), and patient episode initiation for Pap smear (MBS item 73901, an incentive payment for Pap screening). Other component of the screening program cost includes the coordination of the program by the Australian National Cervical Screening Program. The function of the NCSP is to coordinate the program, recruiting participants, monitoring and evaluating quality and standards of pathology services, organisation and management of the Pap Test Register (PTR) within each State/ Territory,

and providing education and promotion of the screening program. The screening program coordination cost is based on the NCSP expenditure from the Department of Health and Ageing and State/Territory governments(Hass, Shanahan et al. 2007). Average coordination and provision costs per woman screened are summarised in Table 3. The coordination cost is less than 10% of total screening program costs and, therefore, we assume the average coordination cost remains unchanged even when the interventions of interest differ in screening intervals and target population.

**Costs associated with abnormal screening results**

The current guidelines *Screening to Prevent Cervical Cancer: Guidelines for the Management of Asymptomatic Women with Screen Detected Abnormalities* were endorsed by the NHMRC in 2005 (National Health and Medical Research Council (NHMRC) 2005), after an extensive literature review and discussions by the multidisciplinary guidelines review group and numerous consultations with clinicians, appropriate professional bodies, and the public. Compared to the previous guidelines, the major changes of the current guidelines are, firstly, the adoption of revised terminology for cervical cytology reporting system known as the Australian Modified Bethesda System (AMBS) 2004; and secondly, a shift to observational management approaches for LSIL based on a greater understanding of the natural history of cervical cancer and HPV infection.

Based on the current NHMRC guidelines, we developed a costing model by an “event pathway analysis” to estimate the costs of cervical cancer and its precancerous lesions. The patient care flow charts in the guidelines are translated into decision trees which are built in EXCEL. Extensive epidemiological data is required to populate the model, predominantly from the results of the national audit studies using the Australian Pap Test Registries(Mitchell, Burrows et al. 2005; National Health and Medical Research Council (NHMRC) 2005), statistical reports from state Cervical Cytology Registry(Mitchell, Burrows et al. 2005), and clinical studies on management of invasive cervical cancer from the literature(Rochelson and Krumholz 1983; Mitchell, Schottenfeld et al. 1998; Roberts, Thurloe et al. 2000; Allen and Narayan 2005). A key assumption around the costing model is that the current NHMRC guidelines and the International Federation Gynaecology and Obstetrics (FIGO) clinical practice guidelines of gynaecologic cancers(Benedet, Bender et al. 2000) are well followed by medical practitioners. Figure 1 provides an example of the costing pathway for HSIL. Similar pathways are developed for LSIL and invasive cervical cancer. Details of the costing methods are available through the author.

Costs are divided into assessment cost, treatment cost, and follow-up costs including the associated costs occurring from the time of an abnormal index cytology report until patients return to routine screening as for the general population with average risk or are dead. Costs are reported as an average cost per each case with abnormal index cytology report. All unit costs for medical and diagnostic service, pharmaceuticals and hospitalisation are derived from Medicare Benefit Schedule (MBS), Schedule of Pharmaceutical Benefits (PBS), and AR-DRG cost weights reported by the National Hospital Cost Database Collection (NHCDC) (National Hospital Cost Data Collection (NHCDC) 2005). A summary of the average cost per abnormal index smears are listed in Table 3.

Figure 1: Pathway analysis for costing HSIL

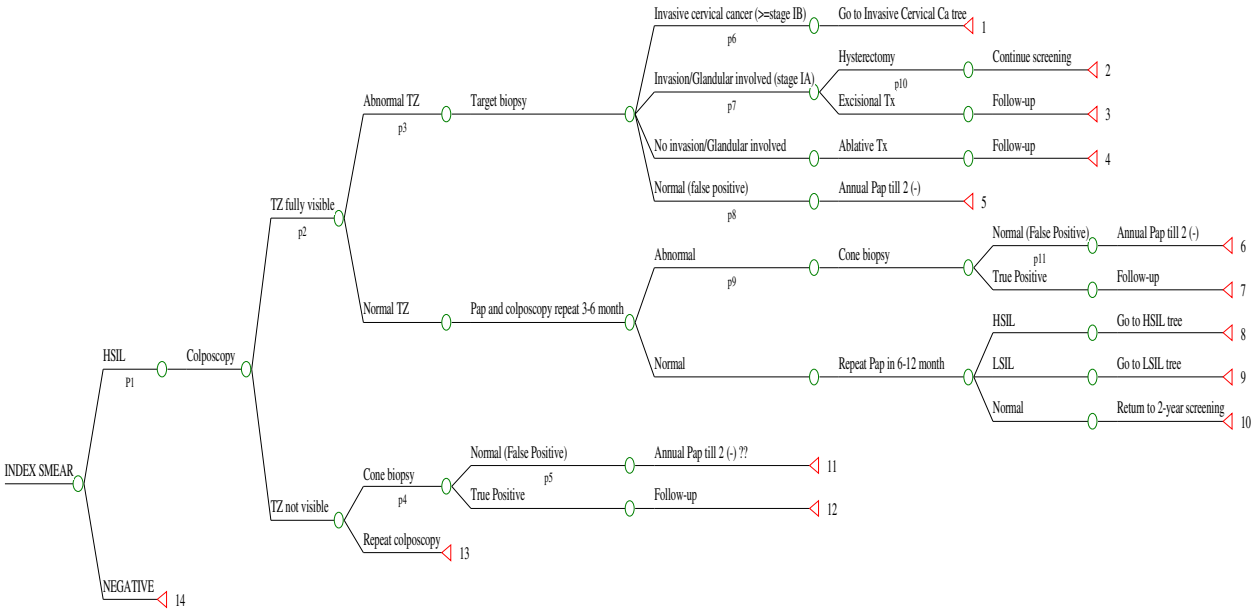



Table 3: Summary of screening cost and cost for each abnormal index smear, not discounted

| **Average screening cost per women screened, in 2003 value** | | | | |
|---|---|---|---|---|
| | **Pap only** | **HPV test only** | **Pap + HPV combined test** | **Source of cost** |
| **Program coordination cost*** | **$9.30** | **$9.30** | **$9.30** | CHERE, 2007 |
| **Screening provision cost** | **$ 85.20** | **$117.45** | **$145.30** | |
| GP Consultation (level C) | $57.35 | $57.35 | $57.35 | MBS, 2003 |
| Screening Pap smear | $19.60 | | $19.60 | MBS, 2003 |

| | | | | |
|---|---|---|---|---|
| Patient episode initiation** | $8.25 | | $8.25 | MBS, 2003 |
| HPV test | | $60.10 | $60.10 | MBS, 2005 |
| **Total average cost** | **$94.50** | **$126.75** | **$154.60** | |

**Average cost of investigation for abnormal index smear**

| | LSIL | HSIL | Micro-invasive | Invasive |
|---|---|---|---|---|
| Assessment Cost | $14,505,371 | $3,650,742 | $26,810 | $459,759 |
| Treatment Cost | $9,601,246 | $12,268,359 | $601,962 | $4,196,057 |
| Follow-up Cost | $8,766,590 | $5,466,883 | $43,868 | $2,556,039 |
| **Total Cost** | **$32,873,207** | **$21,385,984** | **$672,640** | **$7,211,855** |
| No. of women with abnormalities*** | 75,478 | 13,329 | 108 | 570 |
| **Average cost per abnormal index smear** | **$436** | **$1,604** | **$6,239** | **$12,645** |

* Deflated to 2003 value.

** Current incentive payment by Medicare to encourage cervical screening

*** 2003 female age 20-69 without hysterectomy, and participate in screening program.

## 2.7 Sensitivity Analysis

A full uncertainty analysis as per the other ACE-Prevention studies has not been conducted due to the extensive computing requirements to run such an analysis on the cervical cancer stochastic model. One-way sensitivity analysis is carried out to investigate the best and worst scenario on screening tests' accuracy by using the high and low values of 95% CI of the pooled results (refer to Appendix I). Due to lack of data on combined test's sensitivity to detect LSIL as well as sensitivity to detect HSIL for women younger than 30, a threshold analysis was performed to determine the critical point for an ICER to be acceptable by the threshold of $50,000/DALY. Participation rate is modelled with 10% higher and lower than the current level.

Table 4: Uncertainty

| Parameter | Base case | Best/Worst values | Source |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Combined tests - sensitivity for LSIL & HSIL* / specificity | (0.987,0.914) | (1.00,0.918)<br>(0.97, 0.909) | (Petry, Menton et al. 2003)<br>(Cuzick, Szarewski et al. 2003)<br>(Salmeron, Lazcano-Ponce et al. 2003) |
| Combined tests – sensitivity for LSIL | (0.987,0.914) | To be determined by threshold analysis | |
| Participation rate | Age specific participation rates shown in Table 2 | 10% lower,<br>10% higher | Assumption |

* There are no studies reporting sen/spe for combined test using LSIL as a threshold. Therefore, the present analysis assumes same sen/spe for LSIL and HSIL.

# 3. Results

## 3.1 Screening interval

We first evaluated current practice of a Pap smear test with a 2-year screening interval. A marginal analysis was performed to compare the ICERs of 4 different intervals, from 5 to 2 years (Table 5). A 5-year screening interval produces a great health gain and a good cost-effectiveness ratio as compared to no screening. Consecutive shortening of the screening interval produced increasing, but acceptable ICERs all the way down to current practice.

Table 5: Incremental cost-effectiveness ratio of Pap smear for 4 screening intervals

| Screening Interval | Health outcome | | Cost ($m) | ICER |
|---|---|---|---|---|
| Null | Deaths | 6,000 | $202 | |
| | Life-years | 80,000,000 | | |
| | DALY | 73,000,000 | | |
| | | | | |
| 5 years vs. null | Deaths prevented | 2,857 | $595 | $208,000/death |
| | Life-years saved | 56,000 | | $11,000/LY |
| | DALY gained | 47,000 | | $13,000/DALY |
| | | | | |
| 4 years vs. 5 years | Deaths prevented | 405 | $154 | $379,000/death |
| | Life-years saved | 7,000 | | $22,000/LY |
| | DALY gained | 6,200 | | $25,000/DALY |
| | | | | |
| 3 years vs. 4 years | Deaths prevented | 458 | $257 | $560,000/death |
| | Life-years saved | 9,600 | | $27,000/LY |
| | DALY gained | 8,800 | | $29,000/DALY |
| | | | | |
| 2 years vs. 3 years | Deaths prevented | 639 | $513 | $802,000/death |

| | | | | |
|---|---|---|---|---|
| | Life-years saved | 14,000 | | $37,000/LY |
| | DALY gained | 13,000 | | $41,000/DALY |

## 3.2 Null comparison

Next we compare the alternative screening strategies with the null scenario of no screening. A first thing to note is that current practice has already brought huge health gain by preventing more than 4,000 deaths and saving more than 86,000 life years in a cohort of a million women. Any of the alternative screening strategies alter the total amount of health gain by a small margin compared to current practice. All interventions are compared to the null scenario and the results expressed as average cost-effectiveness ratios as outlined in Table 6. The average cost-effectiveness results suggest that three proposed strategies, i.e. HPV test screening every 3 year and combined test screening every 3 years commencing at age 25 or for women over 30 of age, are as good as current practice.

Table 6: Average cost-effectiveness ratio compared to null scenario

| Intervention | Health outcome | | Cost ($m) | C/E |
|---|---|---|---|---|
| Current Practice | Deaths prevented | 4,359 | $1,518 | $348,000/death |
| | Life-years saved | 86,300 | | $18,000/LY |
| | DALY gained | 74,100 | | $20,000/DALY |
| | | | | |
| Strategy 1 | Deaths prevented | 4,282 | $1,383 | $323,000/death |
| Pap screen 2 yearly, | Life-years saved | 83,300 | | $17,000/LY |
| starting at age 25 | DALY gained | 71,600 | | $19,000/DALY |
| | | | | |
| Strategy 2 | Deaths prevented | 4,381 | $1,521 | $347,000/death |
| HPV screen 3 yearly, | Life-years saved | 86,400 | | $18,000/LY |
| current starting age | DALY gained | 75,800 | | $20,000/DALY |
| | | | | |
| Strategy 3 | Deaths prevented | 4,578 | $1,784 | $390,000/death |
| Combined test 3 yearly, | Life-years saved | 90,400 | | $20,000/LY |
| current starting age | DALY gained | 79,900 | | $22,000/DALY |
| | | | | |
| Strategy 4 | Deaths prevented | 4,534 | $1,607 | $354,000/death |
| Combined test 3 yearly, | Life-years saved | 88,100 | | $18,000/LY |
| for women over age 30 | DALY gained | 77,700 | | $21,000/DALY |
| | | | | |
| Strategy 5 | Deaths prevented | 4,559 | $1,617 | $355,000/death |
| Combined test 3 yearly, | Life-years saved | 90,200 | | $18,000/LY |
| starting at age 25 | DALY gained | 79,500 | | $20,000/DALY |

LY: Life-years, DALY: disability-adjusted life years

## 3.3 Incremental analysis compared to current practice

The results of incremental analyses of the alternative screening strategies compared to current practice are presented in a cost-effectiveness plane in Figure 2. The decrement from current practice falls in quadrant III which needs to be carefully interpreted, the greater the ICER the more favourable the intervention is. ICERs in this quadrant that are greater than the threshold (e.g. $50,000/DALY gained) can be considered cost-effective decrements, which means the resources saved can be better spent on other more efficient interventions. Conversely, ICERs less than the threshold in this quadrant indicated that the decrement is not a cost-effective strategy and current practice is to be preferred.

The ICERs for the alternative screening strategies are listed in Table 7.The results suggest that delaying screening commencement age to 25 may be considered on the basis of a cost saving that is marginally better than the threshold. However, there is a considerable health loss, and there are ethical implications of interventions with health loss unless the cost saving could be proven to be used elsewhere with greater health gain.

Four interventions (Strategies 2, 3, 4, and 5) fall in quadrant I where these interventions are more costly for more health gain. In this quadrant, a smaller ICER is more favourable, which means greater health gain with little more cost. These incremental strategies all fall below the ratio of $50,000/DALY (as indicated as the straight line in figure 2), though strategy 2 is the best value for money.

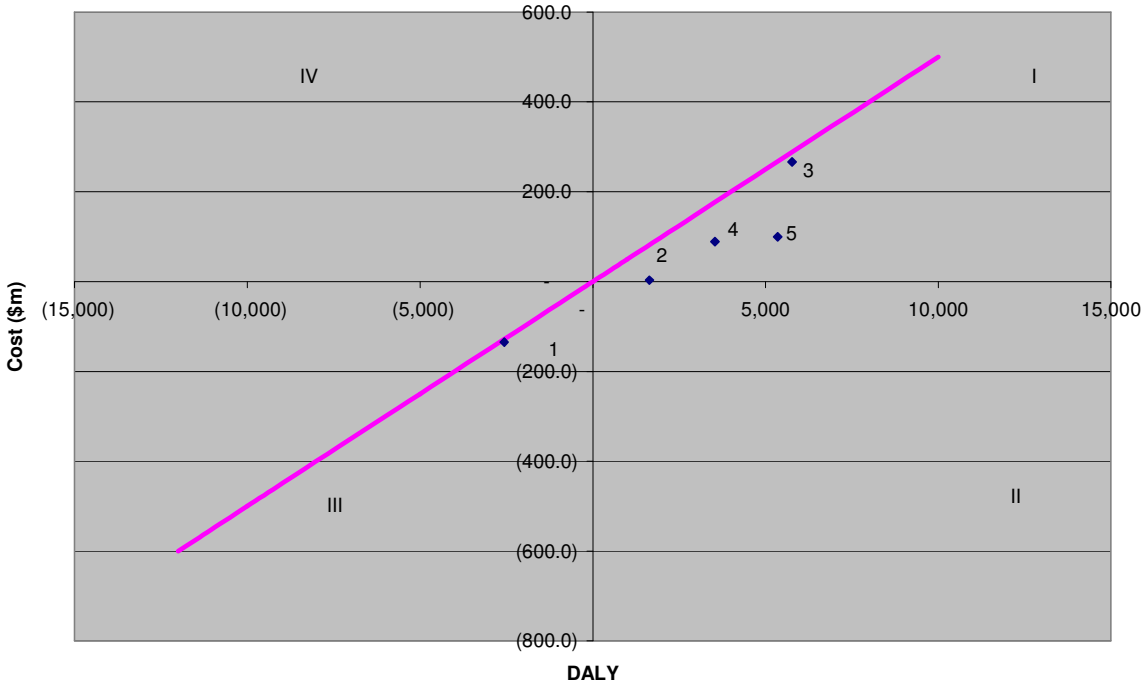
Changing from current practice of Pap smear every 2 year to a combined test every 3 year (strategy 3) is the most expensive option and will add 15% of net cost to the screening program (comparing to total costs in Table 6). However, adoption of combined test for all women screened from age 25 (strategy 5) can achieve equivalent health benefit with only 5% of net cost to the screening program. It was initially thought that applying the combined test to women older than age 30 and continuing use of Pap test every 3 years ages 18 to 30 (strategy 4) might be the most efficient intervention. However, we find that this strategy produces much less health gain at slightly less cost in comparison with using the combined test screening from age 25.

Table 7: Incremental cost-effectiveness ratio compared to current practice

| Intervention | Health outcome | | Cost ($m) | ICER |
|---|---|---|---|---|
| Strategy 1 | Death prevented | -77 | -$135 | $1,750,000/death |
| Pap screen 2 yearly, | Life-years saved | -3016 | | $45,000/LY |
| starting at age 25 | DALY gained | -2565 | | $52,500/DALY |

| | | | | |
|---|---|---|---|---|
| Strategy 2 | Death prevented | 22 | $3.0 | $34,000/death |
| HPV screen 3 yearly, | Life-years saved | 68 | | $13,000/LY |
| current starting age | DALY gained | 1,642 | | $2,000/DALY |
| | | | | |
| Strategy 3 | Death prevented | 219 | $266 | $780,000/death |
| Combined test 3 yearly, | Life-years saved | 4,032 | | $48,000/LY |
| current starting age | DALY gained | 5,766 | | $40,000/DALY |
| | | | | |
| Strategy 4 | Death prevented | 175 | $89 | $312,000/death |
| Combined test 3 yearly, | Life-years saved | 1,803 | | $31,000/LY |
| for women over age 30 | DALY gained | 3,532 | | $22,000/DALY |
| | | | | |
| Strategy 5 | Death prevented | 200 | $100 | $328,000/death |
| Combined test 3 yearly, | Life-years saved | 3,905 | | $16,000/LY |
| starting at age 25 | DALY gained | 5,355 | | $14,000/DALY |
| | | | | |

Figure 2: Incremental cost-effectiveness plane for 5 screening strategies compared to current practice




*Numbers indicate screening strategy number

## 3.4 Sensitivity analysis

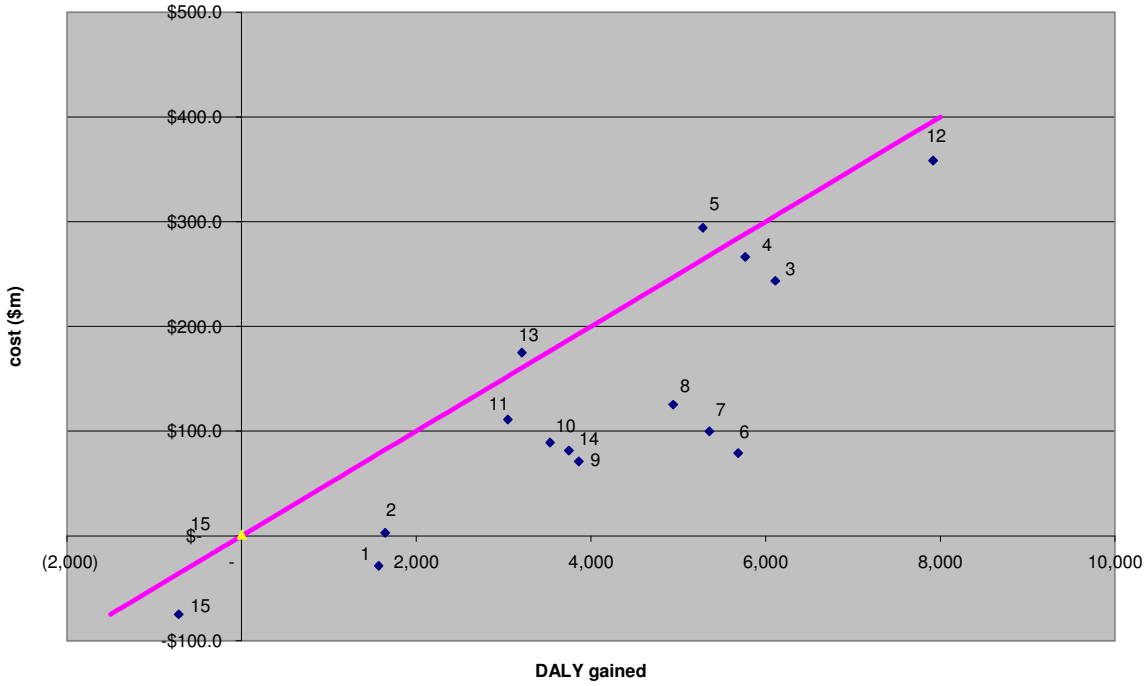
The accuracy of the screening test has a significant impact on the ICERs. The cost-effectiveness result is dominant (i.e. more favourable) by using age-specific estimates of sensitivity and specificity for HPV test compared to current practice (point 1 in Figure 3). In contrast, the ICER is 1,810 per DALY gained by using an overall sensitivity and specificity (point 2). This only strengthens our conclusion that HPV test is more efficient than current Pap smear.

The best/worst scenario and base case for various strategies with combined test are compared in Figure 3. The ICERs for the best, base, and worst for combined test screening every 3 years are $39,866, $46,218, and $55,734 per DALY respectively (point 3, 4, 5). The ICERs of three scenarios for combined test screening every 3 years commencing at age 25 are $13,897, $18,634, and $25,374 per DALY for best, base, and worst, respectively (point 6, 7, 8). Similarly, the ICERs of the best, base and worst scenarios are plotted in point 9, 10, 11, ($18,430, $25,211, $36,430 per DALY) respectively. All results indicate that using combined test to screen women every 3 years commencing at age 25 is the best option for change. Adding the combined test for women aged over 30 is acceptable, but screening all women from age 18 with the combined test may be too expensive.

The threshold analysis has found the critical values of the combined test sensitivity for women younger than 30 and for detection of LSIL to be able to produce an acceptable cost-effectiveness result. For the combined test screening in women aged 18-69 every 3 year, it required a sensitivity of 0.96 to be able to produce an acceptable ICER below the threshold. On the other hand, a sensitivity of 0.86 is needed for adding the combined test in women aged over 30. An even lower sensitivity of 0.80 is necessitated in screening for women aged 25 to 69 every 3 years. A sensitivity of 0.80 (139/173) to detect CIN 1 for HPV test alone has been reported(Cuzick, Szarewski et al. 2003). It is believed a combined test, HPV testing in addition to Pap smear screening, will improve the sensitivity and detect more abnormal cases. Therefore, a sensitivity, ranging from 0.80 to 0.86, of the combined test to detect LSIL and HSIL for women younger than 30 is considered reasonable.

Higher participation improves the ICER for both combined test screen every 3 years (point 12 vs. 13) and HPV test screen every 3 years (point 14 vs. 15). Moreover, if changing the current 2-yearly Pap screen to 3-yearly combined test with a reduced participation rate, produces health gain but with more cost than the $50,000/DALY threshold (point 13). However, it is unclear whether changing from current practice to any proposed intervention has any impact on participation in the screening program; but it is considered unlikely.

Figure 3: Incremental cost and DALY gained compared to current practice from sensitivity analysis



## 3.5 Second stage filter analysis summary

| *Intervention* | *Cost per DALY summary* | *Strength of evidence filter* | *Equity filter* | *Acceptability to stakeholders filter* | *Feasibility & Sustainability filters* | *Potential for side effects filter* |
|---|---|---|---|---|---|---|
| ***Group 1: Age/interval variation from current practice.*** | ***Strategy 1: commencing at age 25 $52,500/DALY (quadrant III) Strategy 1a: 3-yearly $ 40,800/DALY (quadrant III)*** | | ***May have impact on special need population, e.g. in the Indigenous and NESB women.*** | ***Health loss from decrements of current practice is not acceptable by key stakeholders.*** | | ***Negative: Increased cost of litigation due to health loss from decremental interventions (Strategies 1 &1a) (NCCI 2001).*** |
| ***Group 2: HPV test every 3 years*** | ***Strategy 2: $1,800/DALY*** | ***HPV test accuracy is considered sufficient as obtained from a systematic review of all relevant clinical trials.*** | | ***Given the current practice has done a good job; it is hard on political ground to make change.*** | ***HPV test is automated, standardised, and reproducible. Logistically, it is a better test than Pap smear.*** | |
| ***Group 3: Combined test every 3 years*** | ***Strategy 3: 3-yearly $46,000/DALY Strategy 4: adding to women aged over 30 $25,000/DALY Strategy 5: commencing at age 25 $19,000/DALY*** | ***Test accuracy for combined test is a pooled result from 3 studies. Data of combined test to detect LSIL and for women younger than 30 is lacking too.*** | | | | |
| ***Decision point:*** | ***The ICERs are only preliminary and need further work to make a conclusion.*** | ***More work needs to be done, as well as expert advices from cervical cancer epidemiologist, before concluding the strength of evidence.*** | ***How to communicate with the community in particular the special need population needs to be teased out.*** | ***Acceptance of extended screening interval by clinician's and women's health lobby will be an issue.*** | ***With the NCSP in place, altered screening strategies would not greatly impact on the services provision.*** | |

***Policy considerations:***

- ***Current practice has achieved a good cost-effectiveness ratio and variations of screening strategies are marginal adjustments to what have achieved by current practice. HPV testing and combined test appear good value for money for the Australian healthcare system. However, more information on test accuracy and further work with refined model are required to inform health policy.***
- ***There will be immediate impact on cervical cancer screening from HPV vaccination for aged 20-25, as well as for older women who can have HPV DNA testing and vaccinate them if HPV DNA (-).***

***Relevance and suitability for Indigenous Australians:***

- ***HPV test may offer better screening coverage for Indigenous population because of comparable effectiveness from self-collected samples.***

## 4. Discussion

The present analysis has found that current practice already prevents most cervical cancer; options for change of the proposed screening strategies are small adjustments to that have achieved health gain by current practice. The incremental cost-effectiveness of a 2-year screening interval as compared with a 3-year interval is favourable, so extending the screening interval is not advisable. Delaying the screening age to 25 is marginally favourable, but causes a considerable health loss. The adoption of HPV and combined test with an extended screening interval is more costly but affordable, resulting in reasonable ICERs. They appear good value for money for the Australian health care system, but need more information on test accuracy to make an informed decision.

The cost saving of extending the screening interval to 3 years (513 millions) in our analysis is comparable to that reported (50 million for a cohort of 100,000) by Carter (2000) (Carter, Stone et al. 2000). However, the health loss is greater in our analysis in contrast to those in Carter and other studies which have shown little health loss by extending screening intervals (Carter, Stone et al. 2000; van den Akker-van Marle, van Ballegooijen et al. 2002). This is most likely due to the heterogeneity that is a characteristic of our model. Cervical lesions in the model have a distribution of speed of progression that is maintained throughout all the stages. As a result, extending the screening interval produces a large number of interval cancers from fast growing lesions, resulting in a greater health loss. Our study is the first to explore a continuous algorithm for cervical cancer lesions and has a different conclusion in regard to extending the screening interval for Pap smear screening compared to others with deterministic modelling by an average lesion growth. Further research is required to verify this result.

It is important to note that while we endeavoured to use the best quality evidence available there was variation in the literature used to assess the effectiveness of the various strategies. For example, a Canadian study was not included in pooling the combined test accuracy estimate because two thirds of study subjects were under age 35 (refer to Appendix I) and the Canadian study used the Hybrid Capture I (HC-I), less sensitive than later developed Hybrid Capture II as assessed in our study (Ratnam, Franco et al. 2000). In contrast to the Canadian study, the study subjects in other three studies were older than 30 and significant superior test performance was reported (Cuzick, Szarewski et al. 2003; Petry, Menton et al. 2003; Salmeron, Lazcano-Ponce et al. 2003). Therefore, a less accurate combined test reported by the Canadian study is considered to be. On the other

hand, the combined test sensitivity and specificity were corrected for verification bias in the Canadian study; while of the other studies some reported testing random samples of negative cases to ascertain verification bias but seem not to report corrected measures. Therefore, it is difficult to have a head to head comparison between these studies and expert opinion may be warranted.

Although we used a conservative scenario in the sensitivity analysis, the diagnostic accuracy of screening test has a significant impact on the cost-effectiveness results if more extreme sensitivity and specificity is used. The ICERs could swing from favourable (under $20,000 per DALY) to dominated (more cost, less health gain) if the high value of 95% CI of the pooled sensitivity and specificity is used for the combined test as opposed to a extremely worse sensitivity and specificity reported by Ratnam (2000) (Ratnam, Franco et al. 2000) (results not shown).

Clinicians and scientists have been sceptical on the benefit of HPV test in primary screening for cervical cancer(National Cancer Control Initiative 2001). The recently published studies from European large clinical trials truly demonstrate a superior performance of combination of HPV test and Pap smear to detect high grade cervical abnormalities (Cuzick, Szarewski et al. 2003; Petry, Menton et al. 2003) Although the present study over-estimates the sensitivity of combined test for LSIL by assuming the same as HSIL (because no study uses CIN1 as the cut-off threshold), our threshold analysis indicates that a sensitivity of 0.80 to 0.86 is required to produce an acceptable ICER for the combined test screening every 3 years in women aged 25 to 69 or adding the combined test for women aged over 30. Further information and expert opinion needs to be sought to determine whether a better or worse test accuracy than the critical values is likely. Nevertheless, the present study results suggest adding HPV test to current screening is a potential good for value investment to be considered.

The key strength of the current analysis is that we include heterogeneity in our epidemiological model. Each individual modelled has her own propensity to develop lesions, and each lesion gets its own speed of progression determined by a random number drawn from a distribution. This allows some lesions to grow slowly while some grow fast, with the slow growing lesions more likely than the fast growing ones to be picked-up by screening (length bias). This property is rarely built into epidemiological models of cervical cancer (Goldie, Weinstein et al. 1999; Myers, McCrory et al. 2000; Maxwell, Carlson et al. 2002). Participation propensity also differs at the individual level but the average participation rate is still equal to the input average. It is important to differentiate the tendency of participants

and non-participants to participate in a subsequent screening (some non-participants never participate in the program in their life time). As a result, each individual life history is distinct from another. This allows our model to resemble real world scenario better than deterministic models. However, our results on the marginal analysis of screening interval for Pap smear screening have a significant greater health loss compared to earlier studies as discussed earlier, which require further research to validate our stochastic model (Carter, Stone et al. 2000). In addition, the stochastic nature of the model could significantly change the cost-effectiveness results. A full uncertainty analysis is required to make a final conclusion.

Further work will include adding patient time and travel cost and discounting on both cost and health outcomes. More screening strategies and full uncertainty analysis with a multiple life table cohort model will also be developed. Finally, the accuracy of combined test is the most critical input parameter which can greatly impact the cost-effectiveness of the interventions. Therefore, more studies and more accurate estimates will help make better prediction of best screening strategy.

## 5. References

Allen, D. and K. Narayan (2005). "Managing advanced-stage cervical cancer." Best Practice & Research in Clinical Obstetrics & Gynaecology **19**(4): 591-609.

Australian Bureau of Statistics (2006). National Health Survey 2004-5 Summary of Results, Australian Bureau of Statistics, cat. no 4362.0.

Australian Institute of Health and Welfare (AIHW) (1998). Breast and Cervical cancer screening in Australia 1996-1997, AIHW cat. no. CAN 3 Canberra. AIHW. Cancer Series no. 8.

Australian Institute of Health and Welfare (AIHW) (2006). Cervical Screening in Australia 2003-2004, AIHW cat. no. CAN 28 Canberra, AIHW. Cancer Series no. 33.

Bauer, H. M., Y. Ting, et al. (1991). "Genital Human Papillomavirus Infection in Female University-Students as Determined by a Pcr-Based Method." Jama-Journal of the American Medical Association **265**(4): 472-477.

Beggs, S., T. Vos, et al. (2007). The burden of disease and injury in Australia 2003. Canberra:, AIHW.

Benedet, J. L., H. Bender, et al. (2000). "FIGO staging classifications and clinical practice guidelines in the management of gynecologic cancers." International Journal of Gynecology & Obstetrics **70**(2): 209-262.

Bosch, F. X., A. Lorincz, et al. (2002). "The causal relation between human papillomavirus and cervical cancer." Journal of Clinical Pathology **55**(4): 244-265.

Bosch, F. X., M. M. Manos, et al. (1995). "Prevalence of Human Papillomavirus in Cervical-Cancer - a Worldwide Perspective." Journal of the National Cancer Institute **87**(11): 796-802.

Carter, R., C. Stone, et al. (2000). Trial of Program Budgeting and Marginal Analysis (PBMA) to Assist Cancer Control Planning in Australia. Melbourne,, Centre for Health Program Evaluation.

Christine H. Holschneider (2006). Current OB/GYN. Premalignant & Malignant Disorders of the Uterine Cervix, MaGraw-Hill's Access Medicine.

Cuzick, J., C. Clavel, et al. (2006). "Overview of the European and North American studies on HPV testing in primary cervical cancer screening." International Journal of Cancer **119**(5): 1095-1101.

Cuzick, J., A. Szarewski, et al. (2003). "Management of women who test positive for high-risk types of human papillomavirus: the HART study." Lancet **362**(9399): 1871-1876.

Davison, J. M. and J. J. Marty (1994). "Defecting Premalignant Cervical Lesions - Contribution of Screening Colposcopy to Cytology." Journal of Reproductive Medicine **39**(5): 388-392.

Egger, M., G. D. Smith, et al., Eds. (2001). Systematic Review in Health Care. Meta-analysis in context. London, BMJ.

Gaffikin, L., P. D. Blumenthal, et al. (1999). "Visual inspection with acetic acid for cervical-cancer screening: test qualities in a primary-care setting." Lancet **353**(9156): 869-873.

Giles, J. A., E. Hudson, et al. (1988). "Colposcopic Assessment of the Accuracy of Cervical Cytology Screening." British Medical Journal **296**(6629): 1099-1102.

Goldie, S. J., M. C. Weinstein, et al. (1999). "The costs, clinical benefits, and cost-effectiveness of screening for cervical cancer in HIV-infected women." Annals of Internal Medicine **130**(2): 97-+.

Guerra, B., P. De Simone, et al. (1998). "Combined cytology and colposcopy to screen for cervical cancer in pregnancy." Journal of Reproductive Medicine **43**(8): 647-653.

Hass, M., M. Shanahan, et al. (2007). Assessing the costs of organised health programs: the case of the National Cervical Screening Program. Sydney, Centre for Health Economics Research and Evaluation.

Ho, G. Y. F., R. Bierman, et al. (1998). "Natural history of cervicovaginal papillomavirus infection in young women." New England Journal of Medicine **338**(7): 423-428.

Kulasingam, S. L., J. P. Hughes, et al. (2002). "Evaluation of human papillomavirus testing in primary screening for cervical abnormalities - Comparison of sensitivity, specificity, and frequency of referral." Jama-Journal of the American Medical Association **288**(14): 1749-1757.

Loiudice, L., R. Abbiati, et al. (1998). "Improvement of Pap smear sensitivity using a visual adjunctive procedure: a co-operative Italian study on speculoscopy (GISPE)." European Journal of Cancer Prevention **7**(4): 295-304.

Lorincz, A. T. and R. M. Richart (2003). "Human papillomavirus DNA testing as an adjunct to cytology in cervical screening programs." Archives of Pathology & Laboratory Medicine **127**(8): 959-968.

Mann, W., N. Lonky, et al. (1993). "Papanicolaou Smear Screening Augmented by a Magnified Chemiluminescent Exam." International Journal of Gynecology & Obstetrics **43**(3): 289-296.

Maxwell, G. L., J. W. Carlson, et al. (2002). "Costs and effectiveness of alternative strategies for cervical cancer screening in military beneficiaries." Obstetrics and Gynecology **100**(4): 740-748.

Melkert, P. W. J., E. Hopman, et al. (1993). "Prevalence of Hpv in Cytomorphologically Normal Cervical Smears, as Determined by the Polymerase Chain-Reaction, Is Age-Dependent." International Journal of Cancer **53**(6): 919-923.

Mitchell, H., C. Burrows, et al. (2005). Statistical Report 2004, Victoria Cervical Cytology Registry.

Mitchell, M. F., D. Schottenfeld, et al. (1998). "Colposcopy for the diagnosis of squamous intraepithelial lesions: A meta-analysis." Obstetrics and Gynecology **91**(4): 626-631.

Myers, E. R., D. C. McCrory, et al. (2000). "Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis." American Journal of Epidemiology **151**(12): 1158-1171.

Nanda, K., D. C. McCrory, et al. (2000). "Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: A systematic review." Annals of Internal Medicine **132**(10): 810-819.

National Cancer Control Initiative (2001). Moving Forward on Cervical Cytology, Fremantle, WA, National Cancer Control Initiative.

National Health and Medical Research Council (NHMRC) (2005). Screening to Prevent Cervical Cancer: Guidelines for the Management of Asymptomatic Women with Screen Detected Abnormalities, Commonwealth of Australia.

National Hospital Cost Data Collection (NHCDC) (2005). Cost Report Round 8 (2003-2004) AR-DRGv4.2, Commonwealth Department of Health and Ageing,.

Petry, K. U., S. Menton, et al. (2003). "Inclusion of HPV testing in routine cervical cancer screening for women above 29 years in Germany: results for 8466 patients." British Journal of Cancer **88**(10): 1570-1577.

Ratnam, S., E. L. Franco, et al. (2000). "Human papillomavirus testing for primary screening of cervical cancer precursors." Cancer Epidemiology Biomarkers & Prevention **9**(9): 945-951.

Reich, O., H. Pickel, et al. (2001). "Cervical intraepithelial neoplasia III: Long-term outcome after cold-knife conization with clear margins." Obstetrics and Gynecology **97**(3): 428-430.

Roberts, J. M., J. K. Thurloe, et al. (2000). "Subdividing atypical glandular cells of undetermined significance according to the Australian modified Bethesda system - Analysis of outcomes." Cancer Cytopathology **90**(2): 87-95.

Rochelson, B. and B. A. Krumholz (1983). "The Unsatisfactory Colposcopic Examination." Journal of Reproductive Medicine **28**(2): 131-136.

Salmeron, J., E. Lazcano-Ponce, et al. (2003). "Comparison of HPV-based assays with Papanicolaou smears for cervical cancer screening in Morelos State, Mexico." Cancer Causes & Control **14**(6): 505-512.

Sangiva-Lugoma, G., S. Mahmud, et al. (2006). "Visual inspection as a cervical cancer screening method in a primary health care setting in Africa." International Journal of Cancer **119**(6): 1389-1395.

Sankaranarayanan, R., R. Chatterji, et al. (2004). "Accuracy of human papillomavirus testing in primary screening of cervical neoplasia: Results from a multicenter study in India." International Journal of Cancer **112**(2): 341-347.

Tan-Torres Edejer, T., R. Baltussen, et al., Eds. (2003). Making Choices in Health: WHO Guide to Cost-Effectiveness Analysis. Geneva, World Health Organization.

Taylor, S., L. Kuhn, et al. (2006). "Direct comparison of liquid-based and conventional cytology in a South African screening trial." International Journal of Cancer **118**(4): 957-962.

van den Akker-van Marle, M. E., M. van Ballegooijen, et al. (2002). "Cost-effectiveness of cervical cancer screening: Comparison of screening policies." Journal of the National Cancer Institute **94**(3): 193-204.

## Appendix I: Meta-analyses of screening test accuracy

### Conventional Pap test

Nanda et al (2000) reported a systematic review on Pap test accuracy which included 94 studies on conventional Pap test (Nanda, McCrory et al. 2000). However, wide ranges of the reported test performance, e.g. sensitivity ranging from 17% to 99% for threshold of LSIL/CIN1, indicate a need of meta-analysis on Pap test accuracy with more restrict selection criteria. In Nanda's review, 12 studies of the conventional Pap test in screening among low-risk women were selected in which all or a random fraction of patients with negative test results were verified (Giles, Hudson et al. 1988; Mann, Lonky et al. 1993; Davison and Marty 1994; Guerra, De Simone et al. 1998; Loiudice, Abbiati et al. 1998; Gaffikin, Blumenthal et al. 1999). These 12 studies, published from 1992 to 1999, were of good quality by the quality evaluation criteria outlined in the review. Based on the quality evaluation criteria and searching strategy used in the review study, further studies were retrieved from the literature from 2000 to present time. The collection of studies was extended to include another eight studies, which have a number of good quality characters (Sankaranarayanan, Chatterji et al. 2004; Sangiva-Lugoma, Mahmud et al. 2006; Taylor, Kuhn et al. 2006). First, Pap test was used in primary screening as opposed to studies used in clinical follow-up of previous abnormal test result. Second, almost all studies were published in recent 10 years and some were with very large sample size (>10,000) as the data were collected through routine cervical cancer screening program. Third, most studies histology assessments were blinded of the cytology results. Last, all studies have corrected verification bias with some or random sample of negative cases.

Egger 2001 illustrates a simplest method of combining studies' diagnostic accuracy, computing weighted averages of the sensitivity and specificity (Egger, Smith et al. 2001). The estimate of overall sensitivity is dividing the sum of all true positives by the sum of diseased, and similarly the overall specificity is dividing the sum of all true negative by the sum of non-diseased. This method effectively weights each study according to its sample size. Standard error of the estimate is calculated to derive 95% confidence intervals (CI). However, heterogeneity was not investigated by chi-squared test suggested by Egger and therefore, underestimation of test performance may occur by weighted average method if there is an association between the sensitivity and specificity. With the restrict study selection criteria described in the

previous paragraph and Egger's method, pooled sensitivity and specificity of conventional Pap test were calculated for four different cytology and histology thresholds. Results with 95% C.I. are summarised in Table 1.

Table 1: Pooled sensitivity and specificity of conventional Pap test at different threshold.

| Threshold Cytology/Histology | Pooled Sensitivity (95% C.I.) | Pooled Specificity (95% C.I.) | Number of studies included | Number of true positive/diseased | Number of true negative/non-diseased |
|---|---|---|---|---|---|
| LSIL/CIN1* | 0.403 (0.378,0.428) | 0.958 (0.955, 0.960) | 9 | 597/1480 | 17925/18719 |
| LSIL/CIN2+ | 0.628 (0.591,0.666) | 0.963 (0.961,0.965) | 8 | 406/646 | 28629/29735 |
| HSIL/CIN2+ | 0.612 (0.529,0.694) | 0.992 (0.989, 0.994) | 4 | 82/134 | 4723/4762 |
| HSIL/invasive | 1.000 (1.000,1.000) | 0.993 (0.991, 0.994) | 2 | 9/9 | 13904/14008 |

* Sensitivity and specificity were pooled from 7 studies with threshold LSIL/CIN1 and 2 studies with LSIL/CIN1+.

**Human Papillomavirus (HPV) DNA testing**

Cuzick et al (2006) conducted an overview of the HPV DNA testing in primary cervical cancer screening from European and North American countries where routine cytology is in place to provide comparative data (Cuzick, Clavel et al. 2006). The study re-analysed individual patient data for more than 60,000 women by a unified analysis to prevent difficulties in combining data using different entry criteria, cytology threshold and adjustments for verification bias. All studies included in the overview study had a similar split-sample study design and the Hybrid Capture II (HC-II), a commercialized test kit, was used in all studies except one using PCR with GP5+/6+ primer. The pooled test performance results are shown in Table 2.

Because the overview study only reported test performance on histology thresholds of CIN2+ or CIN3+, test sensitivity at threshold of CIN1+ is obtained from two studies which reported data on CIN1+ histology results(Ratnam, Franco et al. 2000; Cuzick, Szarewski et al. 2003). The pooled test sensitivity of histology threshold CIN1+, by the same method (weighted average) used for conventional Pap test, is also included in Table 2.

Table 2: Pooled sensitivity and specificity of HPV DNA testing by age.

| Test performance (Histology threshold) | All | Under 35 | 35-49 | 50+ |
|---|---|---|---|---|
| Sensitivity (CIN2+) | 0.961 | 0.972 | 0.939 | 0.975 |
| Specificity (<CIN) | 0.916 | 0.874 | 0.933 | 0.945 |
| Sensitivity (CIN1+) | 0.713 | N/A | N/A | N/A |
| Specificity (<CIN1) | 0.603 | N/A | N/A | N/A |

**Combined test of Pap and HPV DNA test**

The sensitivity of combined test of Pap and HPV DNA test was reported ranging from 76% to 100%, and specificity ranging from 68% to 95%. The combined test accuracy was age dependent where the sensitivity for women aged over 30 had a significant better result(Lorincz and Richart 2003). In particular, the negative predict value (NPV) for such age group was virtually 100% which warrants its usefulness in screening older women with extended interval. Table 3 summarises the test performance of selected studies with verification on random sample of negative cases. The combined test sensitivity and specificity were pooled using the same method as for Pap and HPV DNA test from three of the studies where published data were available to construct 2x2 tables.

Table 3: Summary of studies on combined test of Pap and HPV DNA testing with histology threshold of CIN2+.

| Study (Country, year) | Threshold*** | Sensitivity (95% C.I.) | Specificity (95% C.I.) | Sample size (age) | Verification |
|---|---|---|---|---|---|
| Pooled | HPV >=1pg, 2pg or cytology >=ASCUS | 0.987 (0.974, 1.00) | 0.914 (0.909, 0.918) | | |
| Cuzick et al* (UK, 2003) | HPV >=2pg or cytology>=mild | 1.00 (0.96,1.00) | 0.94 (0.934,0.945) | 11085 (30-60) | Random 5% sample |
| Petry et al* (Germany, 2003) | HPV >=1pg or cytology >= PapIIw+ | 1.00 (0.937,1.00) | 0.938 (0.918,0.953) | 8466 (>30) | Random 3.4% sample |
| | HPV >=1pg or cytology >= Pap III+ | 1.00 (0.937,1.00) | 0.949 (0.931,0.962) | 8466 (>30) | Random 3.4% sample |
| Ratnam et al** (Canadian, 2000) | HPV >=1pg or cytology >=ASCUS | 0.763 | 0.859 | 2098 (18-69) | Random 10% sample |
| | HPV >=1pg or cytology >=LSIL | 0.763 | 0.893 | 2098 (18-69) | Random 10% sample |
| | HPV >=1pg or cytology >=HSIL | 0.720 | 0.903 | 2098 (18-69) | Random 10% sample |

| | | | | | |
|---|---|---|---|---|---|
| Salmeron et al* (Mexican, 2003) | HPV >=1pg or cytology >=ASCUS | 0.98 | 0.923 | 7868 (15-85) | No verification |

*Diagnosis on histology and colposcopy impression, assuming negative for those without histology and colposcopy

**Diagnosis on histology and colposcopy, corrected estimates

*** Cytology thresholds: mild ~=LSIL, PapIIw~=ASCUS, PapIII+~=LSIL+

# Appendix II: ACE Cervical cancer screen doc

*Jan Barendregt, Wednesday, 2 May 2007*

## *Introduction*

This document serves as documentation for the ACE Cervical cancer screen model, and is a elaboration of the earlier document on the generic ACE cancer screen model, just as the model itself is a derivation of the generic ACE cancer screen model (which is, btw, why the input tables have fields for Males. These can of course safely be ignored). It is not software documentation, but specifies a number of assumptions and algorithms that have been used.
Some general points:

1. The model allows for multiple primary lesions, with a individual propensity to develop one or more.
2. The model makes a distinction between the true cancer status and the clinical status. The former is of course not observed.
3. All cancer events in the following description are conditional on happening before death from all other causes.
4. The model can run two scenarios, one with a user-defined screening programme (see below), the intervention scenario, and without that programme, the null scenario. In both scenarios exactly the same life histories are generated, except for the application of screening. Consequently all the difference in outcomes between the two scenarios is due to the screening programme.

## *Stages*

The model uses a continuous algorithm to simulate the growth of a particular lesion (see the section Algorithm below), but distinguishes along the growth path various stages. These are the following (see the staging proposal in Appendix II for details and justification):

1. No cancer.
2. Low-grade lesion.
3. High-grade lesion.
4. Micro-invasive cancer.
5. Invasive cancer.
6. Distant.

In addition to the true cancer stages above the model distinguishes clinical stages. They have in large part the same names, but a person can be in different true and clinical stages.

1. No cancer.
2. Low-grade lesion.
3. High-grade lesion.
4. Micro-invasive cancer.
5. Invasive cancer.
6. Metastases.
7. Terminal.
8. Disease-free.
9. Recovered.

The Disease-free and Recovered stages are defined to start at a specific time after clinical incidence of (micro-)invasive cancer, Disease-free after a half year (to allow for primary treatment), and Recovered after 5 years. The other clinical stages will often, but not always, be identical to the true cancer stage. In addition to these cancer stages there is a Hysterectomy stage, where no distinction between true and clinical stage is made.
A number of assumptions have been made about the stages:

1. Stages low-grade, high-grade, and micro-invasive are only detected through screening.
2. Only people with true stage Distant proceed to cancer death, but then all of them do (provided death from other causes doesn't come earlier). Consequently, people with true stage Distant at clinical incidence are assigned clinical stages (Micro-)Invasive to allow cancer mortality from those stages.
3. We assume no cancer mortality from clinical stages at incidence of Low- and High-grade.

## *Algorithm*

### Without screening

Note: parameter values given here are in the input table MiscParameters (see also table 1 below), with the names that are given here in *italics*.
Individuals get an individual propensity to develop lesions, expressed as the number of lesions they will develop, given enough time of life. The variable *propensity* is determined by p=round(~LN(1.2,2)) (Note: these parameters are the mean and st dev). With these parameters the mean of p=1.2 (CI95 0  6).
After incidence the tumor is growing using the following equation:

$$S = \exp(\beta t) \qquad 1$$

where $t$ is time (in years) and the variable *growth* $\beta$~LN(0.25,0.2) for women<65 at incidence and $\beta$~LN(0.19,0.15) for women ≥65.
Making the $\beta$ parameter a random draw from a distribution achieves heterogeneity between tumours: some will grow slow, others fast. The age dependency gives women ≥65 on average slower growing tumours, with less heterogeneity.
The symbol $S$ in eq 1 stands for size; however it should not be interpreted literally so. It is not meant to represent physical size (which for some tumours would not make sense anyway), but more as the impact of the tumour on the host, which is assumed to increase exponetially with time.
Eq 1 can be rewritten to get the time until certain tumor sizes:

$$t = \frac{\ln(S)}{\beta} \qquad 2$$

Initially the lesion is in Low-grade stage, the variable *higrade* determines the size at which the tumor goes from Low-grade to High-grade. Currently it is set at 3, which, if beta was not a distribution, would correspond with an average of 4.4 years for women <65, and 5.8 years for women ≥65. Because of the non-linearity the mean of (eq 2) is

larger than that, and more so if the st dev of $\beta$ is larger. For women <65 average duration is 7.2 (CI95 1.4  21.3) years, for women ≥65 9.4 (CI95 1.8  28.8) years.
Low-grade lesions can regress. A time until *lgregression* is drawn from a Weibull, and if this time is shorter than the time to *higrade*, the lesion regresses. The parameters of the Weibull are age-specific, such that the time to regression is short at young ages (mean at age 20: 1.75, CI95 0.3  3.8) and increases with age (mean at age 40: 8.4, CI95 1.6  18.3; at age 85 and over mean: 28.3, CI95 5.1  60.8). Which means that at young age the majority of Low-grade lesions will regress, while at older age few will.
If the low-grade lesion does not regress, it will become high-grade. High-grade lesions can regress as well, using the same mechanism as the Low-grade ones, but with somewhat different parameters that make regression somewhat less likely. The time to regression at age 20 is 2.2, CI95 0.4  4.8, at age 40 11.1, CI95 2.0  24.1, and for 85 and over 31.1, CI95 5.6  67.4.
The total time spend in Low- and High-grade is 16.2 years (CI95 3.2  50.2) for women under 65, and 20.9 years (CI95 4.1  63.6) for women ≥65 ***if no account is taken of the regression.*** With regression these durations will be shorter.
If the High-grade lesion does not regress, it will become micro-invasive at size *microinvasive* (11.7). We assume that no regression takes place anymore once that stage has been reached. The tumour becomes invasive at size *invasive* (17.4), distant at size *distant* (under age 50 at size 21, over 50 19), and causes death at size *death* (32).
Clinical detection takes place between sizes 17.5 and 22 for women <65, and between 18 and 24 for women ≥65. The exact size is determined by taking a draw from a uniform distribution between these sizes. Note that this means that true stage at clinical detection is either invasive or distant. Clinical stage at detection (=incidence) is always invasive. True stage distant gets clinical stage invasive. 15% of the women get a hysterectomy (*hystinvas*).
A diseasefree event is executed *disfree* (put at 0.5) years after clinical detection, and a recover event *recover* (put at 5) years after clinical detection. If tumour size is less than distant, the person will survive the tumour, if more the person will die of the cancer, unless death by all other causes comes first.

| VariableName | Param Name | Lage | Hage | Males | Females |
|---|---|---|---|---|---|
| propensity | par1 | 0 | 111 | 3 | 1.2 |
| propensity | par2 | 0 | 111 | 3 | 2 |
| growth | par1 | 0 | 65 | 0.3 | 0.25 |
| growth | par2 | 0 | 65 | 0.2 | 0.2 |
| growth | par1 | 65 | 111 | 0.3 | 0.19 |
| growth | par2 | 65 | 111 | 0.2 | 0.15 |
| higrade | size | 0 | 111 | 1 | 3 |
| microinvasive | size | 0 | 111 | 3 | 11.7 |
| invasive | size | 0 | 111 | 3 | 17.4 |
| clindetection | par1 | 0 | 65 | 4 | 17.5 |
| clindetection | par2 | 0 | 65 | 7 | 22 |
| clindetection | par1 | 65 | 111 | 4 | 18 |
| clindetection | par2 | 65 | 111 | 7 | 24 |
| distant | size | 0 | 50 | 6 | 21 |
| distant | size | 50 | 111 | 6 | 19 |
| death | size | 0 | 111 | 8 | 32 |

| | | | | | |
|---|---|---|---|---|---|
| disfree | time | 0 | 111 | 0.5 | 0.5 |
| recover | time | 0 | 111 | 5 | 5 |
| metas | size | 0 | 111 | 6.95 | 30 |
| terminal | size | 0 | 111 | 7.8 | 31 |
| hystmicro | percent | 0 | 111 | 0 | 0.48 |
| hystinvas | percent | 0 | 111 | 0 | 0.15 |
| distinvasc | percent | 0 | 111 | 0 | 0.7 |
| lgregression | par1 | 0 | 20 | 6 | 2 |
| lgregression | par2 | 0 | 20 | 2 | 2 |
| lgregression | par1 | 20 | 25 | 2 | 2 |
| lgregression | par2 | 20 | 25 | 2 | 2 |
| lgregression | par1 | 25 | 30 | 2 | 2 |
| lgregression | par2 | 25 | 30 | 2 | 2 |
| lgregression | par1 | 30 | 35 | 2 | 4.5 |
| lgregression | par2 | 30 | 35 | 2 | 2 |
| lgregression | par1 | 35 | 40 | 2 | 7 |
| lgregression | par2 | 35 | 40 | 2 | 2 |
| lgregression | par1 | 40 | 45 | 2 | 9.5 |
| lgregression | par2 | 40 | 45 | 2 | 2 |
| lgregression | par1 | 45 | 50 | 2 | 12 |
| lgregression | par2 | 45 | 50 | 2 | 2 |
| lgregression | par1 | 50 | 55 | 2 | 14.5 |
| lgregression | par2 | 50 | 55 | 2 | 2 |
| lgregression | par1 | 55 | 60 | 2 | 17 |
| lgregression | par2 | 55 | 60 | 2 | 2 |
| lgregression | par1 | 60 | 65 | 2 | 19.5 |
| lgregression | par2 | 60 | 65 | 2 | 2 |
| lgregression | par1 | 65 | 70 | 2 | 22 |
| lgregression | par2 | 65 | 70 | 2 | 2 |
| lgregression | par1 | 70 | 75 | 2 | 24.5 |
| lgregression | par2 | 70 | 75 | 2 | 2 |
| lgregression | par1 | 75 | 80 | 2 | 27 |
| lgregression | par2 | 75 | 80 | 2 | 2 |
| lgregression | par1 | 80 | 85 | 2 | 29.5 |
| lgregression | par2 | 80 | 85 | 2 | 2 |
| lgregression | par1 | 85 | 111 | 2 | 32 |
| lgregression | par2 | 85 | 111 | 2 | 2 |
| hgregression | par1 | 0 | 20 | 6 | 2.5 |
| hgregression | par2 | 0 | 20 | 2 | 2 |
| hgregression | par1 | 20 | 25 | 2 | 2.5 |
| hgregression | par2 | 20 | 25 | 2 | 2 |
| hgregression | par1 | 25 | 30 | 2 | 5 |
| hgregression | par2 | 25 | 30 | 2 | 2 |
| hgregression | par1 | 30 | 35 | 2 | 7.5 |
| hgregression | par2 | 30 | 35 | 2 | 2 |
| hgregression | par1 | 35 | 40 | 2 | 10 |
| hgregression | par2 | 35 | 40 | 2 | 2 |
| hgregression | par1 | 40 | 45 | 2 | 12.5 |
| hgregression | par2 | 40 | 45 | 2 | 2 |
| hgregression | par1 | 45 | 50 | 2 | 15 |
| hgregression | par2 | 45 | 50 | 2 | 2 |
| hgregression | par1 | 50 | 55 | 2 | 17.5 |

| | | | | | |
|---|---|---|---|---|---|
| hgregression | par2 | 50 | 55 | 2 | 2 |
| hgregression | par1 | 55 | 60 | 2 | 20 |
| hgregression | par2 | 55 | 60 | 2 | 2 |
| hgregression | par1 | 60 | 65 | 2 | 22.5 |
| hgregression | par2 | 60 | 65 | 2 | 2 |
| hgregression | par1 | 65 | 70 | 2 | 25 |
| hgregression | par2 | 65 | 70 | 2 | 2 |
| hgregression | par1 | 70 | 75 | 2 | 27.5 |
| hgregression | par2 | 70 | 75 | 2 | 2 |
| hgregression | par1 | 75 | 80 | 2 | 30 |
| hgregression | par2 | 75 | 80 | 2 | 2 |
| hgregression | par1 | 80 | 85 | 2 | 32.5 |
| hgregression | par2 | 80 | 85 | 2 | 2 |
| hgregression | par1 | 85 | 111 | 2 | 35 |
| hgregression | par2 | 85 | 111 | 2 | 2 |

Table 1: content of the input table MiscParameters

A metastatic event is created when the tumour reached size *metas*, diseasefree and recover events are created when they predate the metastatic event. Finally a terminal event is created at size *terminal*, and a death event at size *death*.

## With screening

The screening scenario uses exactly the same life histories as the w/o screening scenario, but adds screening. Screening programmes are defined in the input table Interventions, the content of which is shown in Table 2.

| InterventionName | DiseaseName | Param | Starttime | Stoptime | Lage | Hage | Interval | LGSens | HGSens | MicroSens | InvasSens | DistSens | Specificity | Males | Females |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cervical cancer Screen every 2 year | Cervical cancer | screen | -30 | 110 | 0 | 20 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Cervical cancer Screen every 2 year | Cervical cancer | screen | -30 | 110 | 20 | 25 | 2 | 0.403378378 | 0.628483 | 1 | 1 | 1 | 0.96 | 1 | 0.489 |
| Cervical cancer Screen every 2 year | Cervical cancer | screen | -30 | 110 | 25 | 30 | 2 | 0.403378378 | 0.628483 | 1 | 1 | 1 | 0.96 | 1 | 0.592 |
| Cervical cancer Screen every 2 year | Cervical cancer | screen | -30 | 110 | 30 | 35 | 2 | 0.403378378 | 0.628483 | 1 | 1 | 1 | 0.96 | 1 | 0.6345 |
| Cervical cancer Screen every 2 year | Cervical cancer | screen | -30 | 110 | 35 | 40 | 2 | 0.403378378 | 0.628483 | 1 | 1 | 1 | 0.96 | 1 | 0.641 |
| Cervical cancer Screen every 2 year | Cervical cancer | screen | -30 | 110 | 40 | 45 | 2 | 0.403378378 | 0.628483 | 1 | 1 | 1 | 0.96 | 1 | 0.6425 |
| Cervical cancer Screen every 2 year | Cervical cancer | screen | -30 | 110 | 45 | 50 | 2 | 0.403378378 | 0.628483 | 1 | 1 | 1 | 0.96 | 1 | 0.6565 |
| Cervical cancer Screen every 2 year | Cervical cancer | screen | -30 | 110 | 50 | 55 | 2 | 0.403378378 | 0.628483 | 1 | 1 | 1 | 0.96 | 1 | 0.635 |
| Cervical cancer Screen every 2 year | Cervical cancer | screen | -30 | 110 | 55 | 60 | 2 | 0.403378378 | 0.628483 | 1 | 1 | 1 | 0.96 | 1 | 0.6615 |
| Cervical cancer Screen every 2 year | Cervical cancer | screen | -30 | 110 | 60 | 65 | 2 | 0.403378378 | 0.628483 | 1 | 1 | 1 | 0.96 | 1 | 0.5665 |
| Cervical cancer Screen every 2 year | Cervical cancer | screen | -30 | 110 | 65 | 70 | 2 | 0.403378378 | 0.628483 | 1 | 1 | 1 | 0.96 | 1 | 0.488 |
| Cervical cancer Screen every 2 year | Cervical cancer | screen | -30 | 110 | 70 | 75 | 2 | 0.403378378 | 0.628483 | 1 | 1 | 1 | 0.96 | 1 | 0.181 |
| Cervical cancer Screen every 2 year | Cervical cancer | screen | -30 | 110 | 75 | 80 | 2 | 0.403378378 | 0.628483 | 1 | 1 | 1 | 0.96 | 1 | 0.07 |
| Cervical cancer Screen every 2 year | Cervical cancer | screen | -30 | 110 | 80 | 111 | 2 | 0.403378378 | 0.628483 | 1 | 1 | 1 | 0.96 | 1 | 0.022 |

Table 2: content of the input table Interventions

A screening programme has a unique name, a field for the cancer it applies to, a start and stop time (note that programmes will always be executed in the single cohort population mode, regardless of start time), an age range (note that the age ranges with 0 in the Males and Females fields have no screening), a screen interval (years), sensitivity by cancer stage, specificity, and the Males and Females fields which give average participation.
In table 2 there is the standard programme defined, for the age range 20 and higher, but with a sharply declining participation after age 70. The screen algorithm is as follows:

1. First the person's general propensity *c* to participate is drawn as a uniform random number between 0-1.
2. Age at first screen is drawn as starting age of screen programme + a uniform random number between 0-1.
3. For each screening round participation is determined by drawing a uniform

random number *r* and see whether this is smaller than:

$$\text{if } c > m \quad c + \left(\frac{p-m}{1-m}\right)(1-p)$$

$$\text{otherwise} \quad \left(\frac{c}{m}\right)p$$

3

where *c* is the person's general propensity to participate, *m* is the distribution mean where *r* is drawn from (since this is now a uniform (0-1) distribution *m*=0.5), and *p* is the average participation in the population of that age group. This algorithm ensures that the average participation rate will be equal to the input average, while individuals differ in their propensity to participate, with some attending all rounds and others very few, if any.

4. Age at next round is determined by simply adding the screen interval to the current screen age. This is repeated until the highest age of the screen programme is reached.
5. For each screen round it is first determined whether the person is not a clinical case already, if so the screen is dropped. Next it is determined whether the person has a preclinical cancer. If not a uniform(0-1) random draw that is larger than the specificity of the test will create a false positive event. If the person has a preclinical cancer, a uniform(0-1) random draw that is smaller than the sensitivity of the test creates a true positive event, if larger a false negative event is created.
6. False negative, false positive, and true negative outcomes have no impact on programme flow (although there is an additional cost for false positives, of course).
7. True postive events evolve as follows:
   - First clinical status at incidence is determined. True stages Low- and High-grade, micro-invasive and invasive get the corresponding clinical stages.
   - True stage distant gets clinical stage invasive in 70% of cases (parameter *distinvasc*), and clinical stage micro-invasive in the remaining cases.
   - Of the women in clinical stage micro-invasive, 48% get a hysterectomy (parameter *hystmicro*), of the ones with clinical stage invasive 15% do (*hystinvas*).
   - If not distant, all future cancer events related to this particular instance are removed. If hysterectomy has been carried out, all future cancer events of other instances are removed as well.
   - If clinical stage is micro-invasive or invasive, disease-free and recover events are created starting from current age. Low- and High-grade stages do not generate disease-free and recover events.

## *Output*

Output is substantial, both in number of items as, for some items, the quantity of numbers. Note that, as stated in the introduction, the model distinguishes between the true cancer status and the clinical status, and this distinction means that many of the output items are not observed or even not observable. In the description below I will indicate which output items can be observed empirically (O), and which not (No). Note that this does not mean that the variable is observed in reality, nor that the model output consists of observed data.
The output is available as graphs and files written to disk, and is organised in four sections: Events, Rates, Population, and Check items to write to file.

## Events

This is graphical output to screen, reporting the number of specific events of various kinds in a single year. These numbers depend on the size of the population used in the simulation run. It consists of the following items:

1. Other death. The number of deaths from other causes than cervical cancer. (O).
2. Hysterectomy. The number of hysteroctomies, including those for other reasons than cervical cancer. (O).
3. Low-grade incidence. The number incident cases of low-grade lesions. (No).
4. Low-grade regression. The number regression cases from the low-grade stage. (No).
5. High-grade incidence. The number incident cases of high-grade lesions. (No).
6. High-grade regression. The number regression cases from the high-grade stage. (No).
7. Micro-invasive. The number incident cases of micro-invasive lesions. (No).
8. Invasive. The number incident cases of invasive lesions. (No).
9. Distant. The number incident cases of disseminated disease. (No).
10. Screen. The number of screen events. (O).
11. True positive. The number of true positive screen results. (O).
12. False positive. The number of false positive screen results. (O).
13. False negative. The number of false negative screen results. (No).
14. Clinical detection. The number of cervical cancer cases detected outside the screen programme. These can be either cases in non-participants or interval cancers (O).
15. Disease-free. Number of cases declared disease-free after primary treatment of (micro-)invasive cancer. By definition, someone is declared disease-free when half a year after primary treatment no sign of disease is observed. (O).
16. Recover. Number of cases declared recovered after primary treatment of (micro-)invasive cancer. By definition, someone is declared recovered when five year after primary treatment no sign of disease is observed. (O).
17. Metastasis. The number of cases of clinically detected dissemenated disease. (O).
18. Terminal. The number of cases of clinically defined terminal disease. (O).
19. Death. Deaths from cervical cancer. (O).

Note that several items flagged as (O) here are in reality not or only partially observed. Items 11, 12, and 14, for example.

## Rates

This is graphical output to screen, reporting the number of specific events of as listed above, divided by the number of person years (py) at risk, both in a single year. This immediately poses the question who is at risk? From a theoretical point of view you could argue that, for example, only those women with micro-invasive cancer are at risk to develop invasive cancer. However, since the number of py with micro-invasive cancer is not observed, a rate calculated in that way would have no relation to anything we can observe.

Given that we do not observe py with micro-invasive cancer, nor py with low- and high-grade lesions for that matter, it makes more sense to use py of all women without clinically known cervical cancer, but with a uterus. This still would not be observed,

but at least should show some relation to the clinical incidence of micro-invasive cancer (it will be similar, but higher). On the other hand, the low-grade regression rate, for example, might as well be calculated with low-grade py in the denominator, since no observations on this exists anyway. Below are the definitions of the output rates:

1. Other deaths. Other deaths events, divided by py alive. (O).
2. Hysterectomy. Hysterectomy events, divided by py alive (note that this is a population rate, not a hazard). (O).
3. Low-grade incidence. Low-grade incidence events, divided by py without cancer. (No).
4. Low-grade regression. Low-grade regression events, divided by py Low-grade. (No).
5. High-grade incidence. High-grade incidence events, divided by py Low-grade. (No).
6. High-grade regression. High-grade regression events, divided by py High-grade. (No).
7. Micro-invasive. Micro-invasive events, divided by py w/o clinically known cancer and hysterectomy. (No).
8. Invasive. Invasive events, divided by py w/o clinically known cancer and hysterectomy. (No).
9. Distant. Distant events, divided by py w/o clinically known cancer and hysterectomy. (No).
10. Screen. Screen events, divided by py w/o clinically known cancer and hysterectomy. (O).
11. True positives. True positive events, divided by total number of screen events. Note that this is not a rate but a proportion. (O).
12. False positives. False positive events, divided by total number of screen events. Note that this is not a rate but a proportion. (O).
13. False negatives. False negative events, divided by total number of screen events. Note that this is not a rate but a proportion. (No).
14. Clinical detection. Clinical detection events, divided by py w/o clinically known cancer and hysterectomy. (O).
15. Disease-free. The proportion of clinical incident cancer cases that is followed by a disease-free event. (O).
16. Recover. The proportion of clinical incident cancer cases that is followed by a recover event. (O).
17. Metastasis. The proportion of clinical incident cancer cases that is followed by a metastasis event. (O).
18. Terminal. The proportion of clinical incident cancer cases that is followed by a terminal event. (O).
19. Case fatality. The proportion of clinical incident cancer cases that is followed by a cancer death event. (O).

## Population

This is graphical output to screen of a variety of output items.

1. Population. Number of people by 1 year age groups and ‘true’ cancer status. (No).

2. Person years at risk. Person years by the chosen age group schedule and 'true' cancer status. (No).
3. Clinical status. Person years by the chosen age group schedule and clinical status. (O).
4. Lead time. Distribution of lead times by age group. (No).
5. Pre-clinical stage durations. Distribution of pre-clinical stage durations by age group. (No).
6. Mortality. The mortality rate of cervical cancer, by age group, calculated by number of deaths events, divided by py alive. (O).
7. Clinical incidence. The clinical incidence rate, including low- and high-grade lesions, calculated as the sum of clinical detection and true-positive events, divided by py w/o clinical cancer. (O).
8. (Micro-)invasive. The clinical cancer incidence rate, calculated as the sum of screen- and not-screen-detected micro-invasive, invasive, and distant cases at incidence , divided by py alive (note that this is a population rate, not a hazard). (O).

### Check items to write to file

This section describes the items that can be written to a comma-delimited file. The items mostly consist of those described in the three output sections above.

1. Population. This is item 1 of the Population section.
2. Event numbers. All items in the Events section.
3. Person years at risk. Item 2 of the Population section.
4. Event rates. All items in the Rates section.
5. Life histories. The individual life histories that form the basis of all output. This item is enabled only when the 'Store life histories' checkbox on the main window has been enabled before the calculation. Note that enabling this checkbox greatly increases memory use, and also note that the resulting life history output can be huge, depending on numbers of persons and years simulated. (No).
6. Stage distribution. The 'true' cancer stage at clinical incidence, by detection mode (screen or not). (No).
7. Clinical status person years. Item 3 of the Population section.
8. Population rates. Items 6, 7 and 8 of the Population section.
9. Clinical stage distribution. The numbers of incident cases by age and stage as it is assessed at clinical incidence, by detection mode (screen or not). (O).
10. Lead time. Item 4 of the Population section.
11. Pre-clinical stage durations. Item 5 of the Population section.
12. Survival by age and clinical stage at incidence. (O).

## *Fitting the model*

Fitting the model means adjusting the parameters such that the model creates unobserved individual life histories that together add up to the observed population data, and confirm to common sense and some notions about the pre-clinical disease natural history. This is a rather indirect and fiddly process.

Available data are the current incidence and mortality of cervical cancer in Australia. Since these are observed in a population with a screening program in place, I had to

create screening program input that describes the current situation in terms of age range, participation, interval, and sensitivity and spcificity of the Pap-smear test. Then the intervention scenario of the model was fitted to the observed incidence and mortality data.
There is also some data from the Victorian Cancer Registry on clinical stage at incidence by age, but it soon turned out to be inconsistent with the mortality data. With that stage distribution the model produced a far too low mortality, so in all likelihood the registry data are from a selected group. Consequently this data has not been used for fitting.
Fitting was done using the single cohort population option of the model. This implies that past trends were ignored during the process.

## *Results*

### Incidence

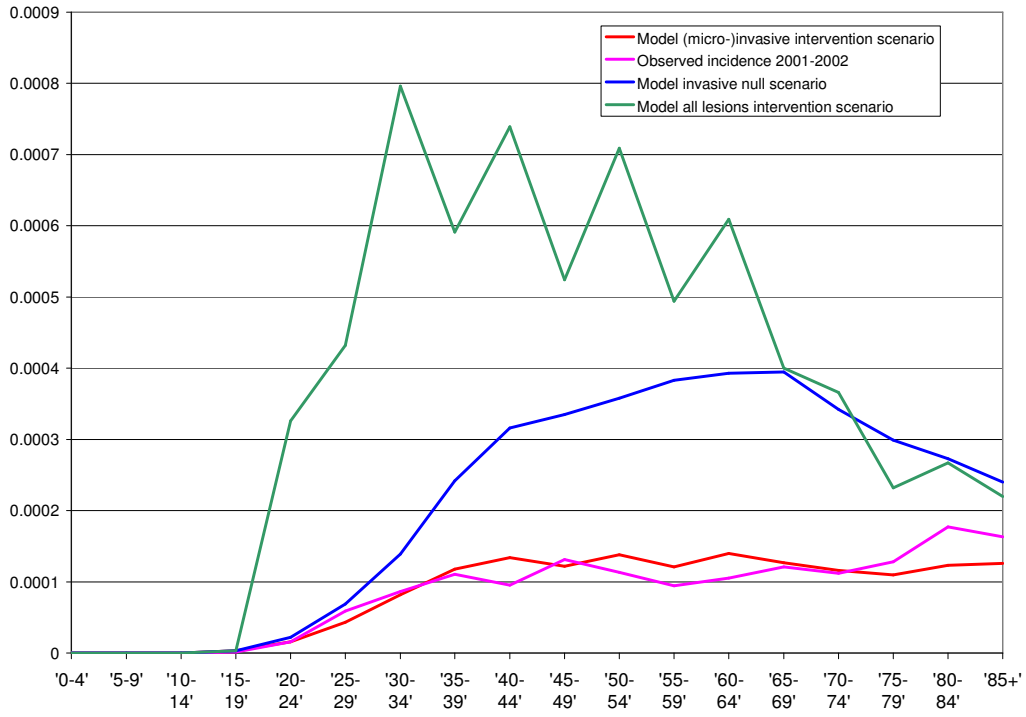



Figure 1 shows various kinds of incidence by age. At the bottom are the observed incidence of micro-invasive and invasive cancer (averaged over 2001 and 2002) and fitted model incidence from the intervention scenario. For comparison the figure also shows the incidence of invasive cancer in the absence of screening: clearly a large part of cancer incidence is prevented by the screening program. The drawback can be seen from the incidence of all lesions (so including Low- and High-grade): about 5-6 times as high as cancer incidence proper. The sawtooth effect, btw, is due to the two-year screen interval, which causes in some 5 year age groups three rounds, in others two.

## Mortality

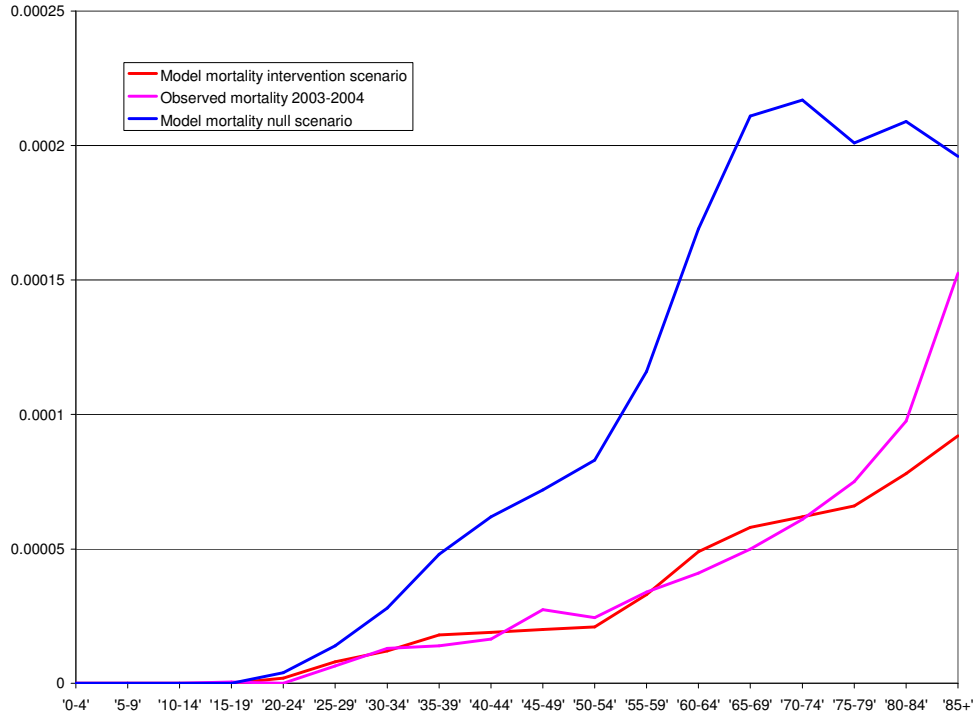
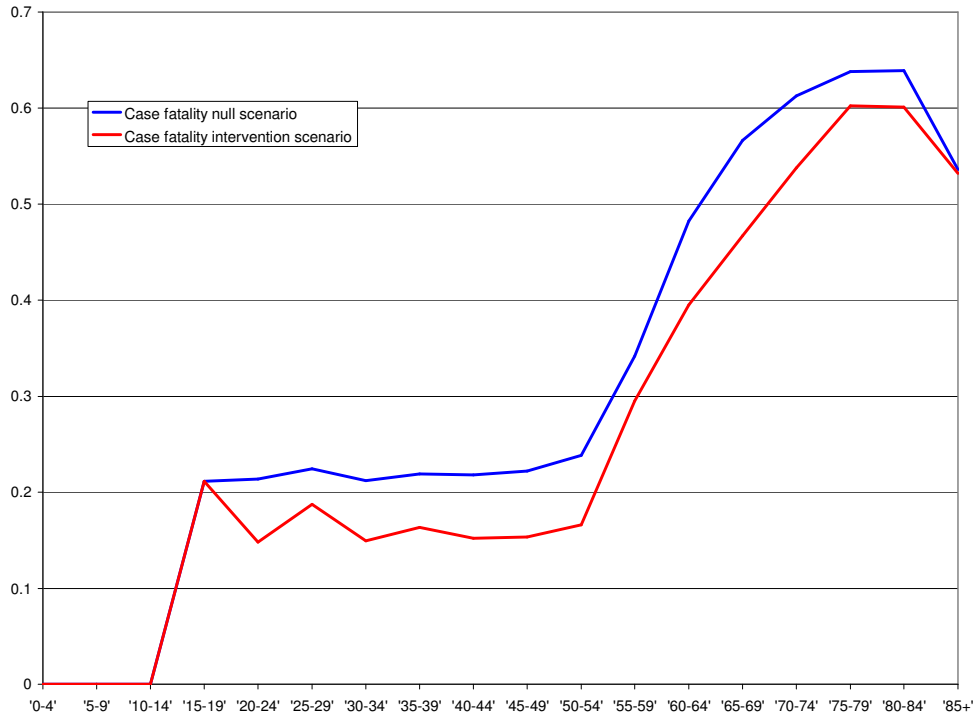



Figure 2 shows the fitted model mortality from the intervention scenario, plus the observed mortality (averaged over 2003-2004). Again the results from the null scenario are also shown: clearly the screening program prevents considerable mortality. This is due not just to the lower incidence, but also to a lower case fatality: in the intervention scenario the screen-detected cancers have an earlier stage. The effect on case fatality is shown in Figure 3:

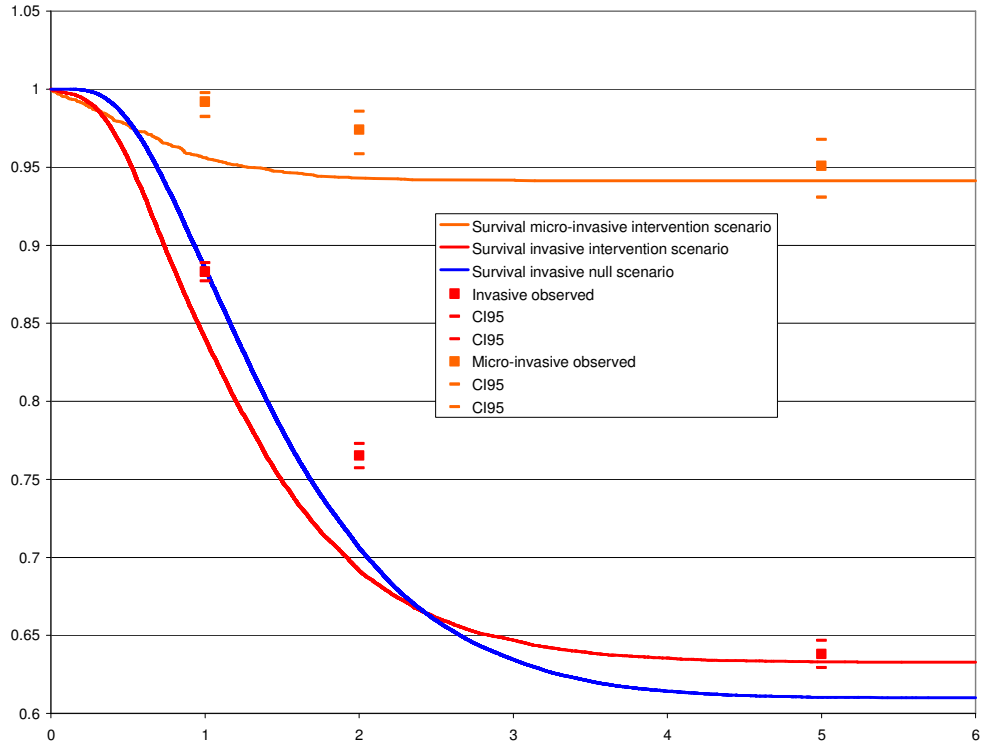


## Survival

Survival data by stage, including 95% CIs, is in Table 3.

| | | | | | |
|---|---|---|---|---|---|
| Micro-invasive | 518 | survivors | proportion | lower CI | higher CI |
| | 1 year | 514 | 0.992 | 0.983 | 0.998 |
| | 2 year | 505 | 0.974 | 0.959 | 0.986 |
| | 5 year | 493 | 0.951 | 0.931 | 0.968 |
| Invasive | 11427 | | | | |
| | 1 year | 10091 | 0.883 | 0.877 | 0.889 |
| | 2 year | 8744 | 0.765 | 0.757 | 0.773 |
| | 5 year | 7292 | 0.638 | 0.629 | 0.647 |

Figure 4 gives the results from the model for all ages, for micro-invasive and invasive from the intervention scenario plus observed survival and the 95% confidence intervals, and for invasive from the null scenario.




Generally the model seems to let the women die from their cancer a bit sooner than the observed data, but the 5-year survival is well within the 95% CI. The age pattern is that survival is better for younger age groups.

## Stage durations

Table 4 shows average stage durations (years) by age. The general pattern is that stages last longer for older women, are longer in the null than in the intervention scenario, and are longer for earlier stages.

| | Intervention scenario | | | | Null scenario | | | |
|---|---|---|---|---|---|---|---|---|
| Age | Low-grade | High-grade | Micro-invasive | Invasive | Low-grade | High-grade | Micro-invasive | Invasive |
| '10-14' | 1.47 | 1.53 | 0.35 | 0.08 | 1.47 | 1.53 | 0.35 | 0.08 |
| '15-19' | 1.47 | 1.24 | 0.38 | 0.11 | 1.47 | 1.24 | 0.38 | 0.11 |
| '20-24' | 1.36 | 1.24 | 0.45 | 0.14 | 1.54 | 1.50 | 0.49 | 0.15 |
| '25-29' | 1.46 | 1.52 | 0.49 | 0.15 | 1.64 | 1.86 | 0.54 | 0.16 |
| '30-34' | 1.75 | 1.56 | 0.54 | 0.17 | 2.12 | 2.18 | 0.64 | 0.19 |
| '35-39' | 2.87 | 2.06 | 0.67 | 0.20 | 3.58 | 2.76 | 0.78 | 0.22 |
| '40-44' | 2.87 | 2.50 | 0.83 | 0.26 | 4.42 | 3.73 | 1.05 | 0.31 |
| '45-49' | 3.25 | 2.85 | 0.94 | 0.29 | 5.16 | 4.17 | 1.19 | 0.36 |
| '50-54' | 3.22 | 2.83 | 0.99 | 0.31 | 5.87 | 4.58 | 1.31 | 0.39 |
| '55-59' | 3.55 | 3.30 | 1.08 | 0.31 | 6.63 | 5.07 | 1.43 | 0.40 |
| '60-64' | 3.66 | 3.34 | 1.16 | 0.30 | 7.47 | 5.50 | 1.56 | 0.39 |
| '65-69' | 4.27 | 4.00 | 1.29 | 0.29 | 8.11 | 5.97 | 1.70 | 0.39 |
| '70-74' | 5.03 | 4.61 | 1.45 | 0.30 | 9.07 | 6.72 | 1.89 | 0.39 |
| '75-79' | 7.38 | 5.63 | 1.73 | 0.35 | 10.15 | 7.34 | 2.19 | 0.44 |
| '80-84' | 9.44 | 6.09 | 1.89 | 0.45 | 12.71 | 7.98 | 2.49 | 0.56 |
| '85+' | 13.89 | 7.51 | 2.76 | 0.60 | 16.21 | 9.24 | 3.28 | 0.66 |

## Lead time

The whole purpose of screening is producing lead time: the length of time a lesion is detected earlier as a consequence of the screening. Table 5 shows average and median lead time produced by the current screening program. The average is higher than the median, evidence of a skewed distribution with some very large values. The age pattern is for short lead times at younger ages (because of short stage durations), increasing into middle age (as the stage durations increase), only to decrease again into old age (as remaining life span and screening participation decreases).

| | Lead time | |
|---|---|---|
| Age | Average | Median |
| '20-24' | 3.01 | 2.52 |
| '25-29' | 3.13 | 2.70 |
| '30-34' | 5.40 | 4.62 |
| '35-39' | 5.99 | 5.03 |
| '40-44' | 7.29 | 6.02 |
| '45-49' | 8.03 | 6.64 |
| '50-54' | 8.49 | 7.15 |
| '55-59' | 8.58 | 7.29 |
| '60-64' | 8.11 | 6.92 |
| '65-69' | 7.71 | 6.59 |
| '70-74' | 7.28 | 6.40 |
| '75-79' | 6.19 | 5.47 |
| '80-84' | 4.94 | 4.27 |
| '85+' | 3.54 | 2.52 |

## *Discussion*

The present model would benefit from scrutiny by experts in the field of cancer screening in general, and Australian cervical cancer screening in particular. There are a lot of assumptions in the model, and, since we model an unobserved pre-clinical

natural history, data is by definition limited, and expert knowledge indispensable. A few observations:

1. The general patterns seem OK. The model does a reasonable job reproducing observed data (incidence, mortality, and survival), the outcomes of unobserved data also seem reasonable (lead time, stage durations), and outcomes change in expected directions when different scenarios are calculated.
2. Nevertheless it would be good to have a more definite data set to fit the model to. Cervical cancer is a rare disease, and there are surprisingly large differences in incidence and mortality between years. Also additional data, such as stage distribution at incidence, would be very valuable.
3. In addition, statistical analysis of the simulated individual life histories would be useful, if only to find any anomalies.
4. There is considerable heterogeneity between the simulated life histories. In particular, the growth speed of the lesions, governed by the β parameter of eq 1, is very different, with some lesions speeding through the various stages, while others are, for all practical purposes, non-progressing. This heterogeneity causes the average lead time to be higher than the median. It is probably also the cause of the absence of lead time bias in the survival curves: screening picks up slow-growing lesions sooner than fast-growing ones (length bias), therefore with a screening program in place the lesions that reach the (micro-)invasive stage are selected to be fast-growing ones.
5. The fact that the modelled 5-year survival is the same as the observed, while the 1- and 2-year survivals are worse, might indicate that the degree of heterogeneity is too large. But it might also be the case that the treatment temporarily slows down cancer progression, and this effect is currently not modelled.

I plan to write an article that reports the results of the model in comparison with the null scenario, as estimates of what health gain and losses the current screening program produces.

**Appendix III: Proposal for staging cervical cancer in ACE-Prevention**

The purpose of this document is for cancer experts to provide comments to the researchers on the ACE-Prevention project[1] regarding a staging system – detailed below – proposed for use in the development of a cancer screening model (this model will be used to assess the cost-effectiveness of different screening strategies, such as reducing or increasing the screening interval). The staging should minimally distinguish cervical abnormalities by differences in treatment and survival.

## Background:

There are two reporting systems related to the staging of cervical abnormalities, cytological (Pap results) and histopathological (biopsy results). Histopathological results are uniformly reported in Cervical Intraepithelia Neoplasia (CIN) terminology. Unfortunately, cytological results have been reported by different terminological systems. Although the Bethesda system for reporting Pap smears has been widely used internationally, Australia has developed its own unique terminology system. In 1994, the National Health and Medical Research Council prepared the first guidelines for the management of women with screen-detected abnormalities. The associated working party considered the Bethesda terminology and recommended a range of modifications which resulted in the NHMRC endorsed Australian terminology. In 2004, the NHMRC introduced new guidelines for the management of asymptomatic women with screen detected abnormalities and Australia adopted a revised terminology system for cervical cytology, known as the Australian Modified Bethesda System 2004 (AMBS 2004).

## Proposed Staging:

At this stage we believe that the incidence of cervical cancer in the ACE-Prevention cancer screening model should be based on histopathological terminology. Histopathological testing presents confirmed status of cervical abnormalities by biopsy tissue samples. In contrast, cytological testing is only a predictor of the disease and requires histological confirmation. Current histological findings are defined by the Cervical Cytology Registry as the follows:

[1] ACE-Prevention is a 5 year NH&MRC funded research project with the aim of assessing the cost-effectiveness of 150 preventative interventions for non-communicable disease in order to inform Australian health policy. The top three chief investigators of the project are Associate Professors Theo Vos, Rob Carter and Chris Doran.

- Invasive cancer
- Microinvasive cancer
- CIN 3 with questionable microinvasion
- CIN 3
- CIN2/3
- CIN 2
- High grade – not otherwise defined
- CIN – not otherwise defined
- CIN 1
- HPV effect
- Low grade – not otherwise defined
- Benign changes
- Normal

For modelling purposes, the above grading are too numerous to be useful. For example, follow-up does not differ for women with HPV effect and CIN 1; and treatment plans are identical for CIN 2 and CIN 3. Therefore, detailed grading (e.g. CIN 2, CIN 2/3, CIN 3) does not offer benefit but create unnecessary complexity for modelling. We therefore propose five groups of staging, set out in Table 1. The rationale of this 5-groups staging is based on treatment and management of outcomes. In the new NHMRC (2005) guidelines for the management of asymptomatic women with screen detected abnormalities, treatment and management of low-grade & high-grade squamous abnormalities and glandular abnormalities are clearly described. The constitution of each group (low-grade squamous abnormality, high-grade squamous abnormality, and glandular abnormalities) is also clearly defined in the guidelines. Using this grouping definition, our 5-group staging corresponds very well to those defined in the guidelines for the purpose of treatment and management. In terms of survival, there is no difference in mortality between low-grade and high-grade intraepithelial abnormalities although high-grade has a higher risk of recurrence and progression to invasive cancer (Mitchell H & Hocking J, 2002). However, there is a difference of survival between micro-invasive and more advanced invasive cervical cancer (refer to Table 1).

Secondly, as ACE-Prevention aims to inform Australian health policy, we prefer to use Australian data sources if available. Many Australian studies report cervical abnormalities based on histological testing as high-grade and low-grade which directly correspond to the staging levels we suggest in table 1. Examples of such

studies include Mitchell (2005)[2] and Mitchell and Hocking J (2002)[3]. The National Cervical Screening Program has reported age-specific incidence rates of micro-invasive and all cervical cancers in the community with data back to 1983 and onward.

[2] This national audit study used data provided by the Australian Pap test registries to investigate the outcomes after a cervical cytology report of low-grade abnormality. The study included 76,709 women with index smear of low-grade abnormality who had further information available during the 24 month follow-up. The cross-sectional histology results at 6 month follow-up and the longitudinal outcomes over the 24-month period were presented

[3] This cohort study determined the risk of recurrent abnormality after a first episode of high-grade epithelial abnormality and its evolution over time in a population setting. The rate ratios of subsequent high-grade epithelial abnormality and subsequent invasive cancer were reported for three age groups.

Table 1: Corresponding staging between CCR and ACE-Prevention modelling, and treatment procedures & survival for each stage

| Cervical Cytology Registry Staging | ACE-Prevention Staging | Treatment | Follow-up | Survival |
|---|---|---|---|---|
| Invasive cancer<br>Adenocarcinoma | Invasive cancer | - Radical abdominal hysterectomy & pelvic lymphadenectomy<br>- External radiation & intracavitary brachytherapy with/without concurrent chemotherapy<br>- Systemic chemotherapy<br>- Pelvic exenteration | Pelvic examination every 3-4 months in the first 2 years after completing treatment and 6 monthly thereafter. Yearly follow-up is appropriate after 5 years following treatment.[4] | 1 year -88%<br>2 years -77%<br>5 years -64% |
| Micro-invasive cancer<br>CIN 3 with questionable micro-invasion | Micro-invasive cancer | - Hysterectomy<br>- Excisional therapy (e.g. cold knife cone biopsy, LEEP) for women with fertility concern | Same as high-grade intraepithelial disease | 1 year -99%<br>2 years -97%<br>5 years -95% |
| CIN 3<br>CIN 2/3<br>CIN 2<br>High grade – not otherwise defined<br>CIN – not otherwise defined<br>Adenocarcinoma in situ (AIS)<br>Possible high grade lesion | High-grade intraepithelial lesion | - Ablative therapy (e.g. CO2 laser) for high-grade without evidence of invasion and glandular lesion<br>- Excisional therapy (e.g. cold knife cone biopsy, LEEP) for high-grade suspicion of early invasion & AIS, and unsatisfactory colposcopic assessment | Colposcopy and cervical cytology at 4-6 months after treatment. Cervical cytology and HPV typing should then be carried out at 12 months after treatment and annually thereafter until the women has tested negative by both tests on 2 consecutive occasions; and then return to routine screening as the average population. | Assuming 5-year survival of high-grade intraepithelial disease is 100%, given 99.65% of CIN III with clear margins on cone biopsy remaining free of disease after a mean follow-up of 18 years (Reich et al, 2001) and 78% of CIN III with involved margins on cone biopsy remaining free of disease during a mean follow-up of 19 years. (Reich et al, 2002). (CIN III clear margin : involved margin = 82% : 18%)<br>Australian data suggested a crude rate of cervical cancer in CIN 2&3 was 0.35 per 1000 person years. (Heather, M & Hocking J 2002) |

[4] Allen D. & Narayan K. (2005) Managing advanced-stage cervical cancer. *Best Practice & Research Clinical Obstetrics and Gynaecology* 19(4): 591-609

| | | | | |
|---|---|---|---|---|
| CIN 1<br>HPV effect<br>Low grade – not otherwise defined<br>Atypical endocervical or glandular cells of undetermined significance | Low-grade intraepithelial lesion | Repeat Pap test in 6-12 month | Return to routine screening as the average population after 2 consecutive annual Pap tests showing negative. | Assuming 5-year survival of low grade intraepithelial disease is 100%. |
| Benign changes | Normal or benign | | | |