A simulation model to explore the cost-effectiveness of policy changes to the UK cervical screening programme

Pelham Barton
Stirling Bryan

Health Economics Facility
University of Birmingham

Correspondence:

Pelham Barton
HSMC
Park House
40 Edgbaston Park Road
BIRMINGHAM
B15 2RT

Acknowledgment:

This work is part of a project funded through the National Co-ordinating Centre for Health Technology Assessment: Cervical screening programmes: can automation help? (Project code: 98/38/01). The main project worker was Dr Brian Willis and the project was co-ordinated by Dr Chris Hyde, both from the Department of Public Health and Epidemiology in the University of Birmingham. The description of the screening programme was mainly due to Dr Willis and to Dr Philippa Pearmain of the West Midlands Cancer Intelligence Unit.

The authors of this paper take full responsibility for all statements made here.

Background

Cervical cancer, its detection and treatment, remains an important area of activity for the NHS[1]. However, cervical cancer is a relatively rare cause of death in the UK (crude annual death rate 4.5 per 100,000[2]) and it is mass screening which has heightened the profile of the disease. This screening has a strong rationale. It usually takes eleven to twelve years for cervical dysplasia to progress to invasive neoplasia[3 4], so offering the opportunity to detect and treat the disease in the pre-invasive stage. Before the introduction of screening for cervical cancer in the 1950s, there were over 2,500 deaths per year in England and Wales; now there are around 1,200[2].

Recently the strain on the NHS Cervical Screening Progamme (NHSCSP) has become acute, due to a combination of media attention, pressure to reduce screening intervals and difficulty recruiting and retaining screeners. New technologies appear to offer solutions to some of these pressures. An example is the automation of slide analysis in which slides are translated into computerised images, subject to analysis by computers and a categorisation into normal or abnormal made. Automated image analysis is not currently intended to replace the manual system relying on human interpretation of slides through a microscope, but rather to work with it. Implementation costs associated with automated devices would be considerable: £40 million is a very conservative estimate for the NHSCSP. The use of technologies other than automated image analysis, such as liquid based cytology (LBC) and human papillomavirus (HPV) screening are also being actively considered by many cervical screening programmes, including the NHSCSP. In addition, the role of non-technological approaches to improving cervical screening, particularly those with a focus on increasing coverage, should not be overlooked. The wide range of policy questions currently being considered in the delivery of cervical screening highlights the need to systematically consider alternative options within a modelling framework.

The focus for the work reported in this paper was the issue of automation (and in particular the introduction of the AutoPAP technology). The paper briefly reviews previous modelling work undertaken in the area of cervical screening, before describing

in detail the model we have constructed and the data we have drawn on. The limitations of our work to date are then discussed.

Previous Modelling

The existing models are essentially derived from three sources[5 7 8]. The AHCPR model by McCrory *et al*[5] and the Payne *et al*[7] models are two independent Markov models. The main difference between them is that AHCPR includes HPV and Payne *et al* does not. These models operate in a fixed cycle (yearly for AHCPR and 6-monthly for Payne *et al*). Each model assumes that screening occurs at fixed ages and does not allow for surveillance screening. Payne *et al* allows for a small number of inadequate smears, but assumes immediate re-testing; "subsequent slides are assumed to be adequate".

Cuzick *et al*[8] use a semi-Markov approach; this model has the advantage that it allows individual patient histories to be followed and includes surveillance screening. This model does not seem to allow for inadequate smears, although its structure would presumably accommodate them without much difficulty.

None of the models takes account of delays in screening results caused by queuing effects. The only way this can be done is to use a simulation methodology which allows for interaction between individuals. One such methodology is discrete event simulation (DES).

A population-based DES model tracks individuals (in this case women) from an entry point (in this case, reaching the age of 15 years) to an exit point (in this case, all women are followed through to death). Women may change state at various times through the model; it is assumed that such changes take no time. Pseudo-random numbers are used where appropriate to generate the necessary variations within the model. These relate both to time spent in a given state, and pathway followed between states. The important feature of DES that is relevant to this paper is that it allows for the effects of limited resources. In this case if the number of smears sent to the laboratory over a period of time exceeds the capacity to process the smears, then the turn-around time will increase until the rate of receipt of new smears at the laboratory reduced below the handling

capacity. Note that in this model, women of different ages are screened at the same calendar time, and so are competing for resources in a realistic way. A further relevant feature of DES modelling is that it is possible to assess the transient effect of a change in policy as well as the new steady state to be reached.

The Model Structure

Our model is a discrete event simulation (DES) model, written in Borland Delphi, using an event-based executive. Individual women pass through the model, their status being changed at various times according to the natural history of cervical cancer and the effects of a screening programme. The natural history sector of the model is based on that used by McCrory *et al* (AHCPR)[5], while the screening procedure is based on a specially developed patient walkthrough (available from authors on request). The starting point for this was a translation of the NHSCSP guidance[6] into a step-by-step algorithm. Areas of uncertainty were resolved by discussion with appropriate experts.

AHCPR[5] reports a Markov model with a one-year cycle which follows a cohort of women from age 15 to age 85. For the model reported here, this has been converted into a DES model. The key differences in model structure are:

- Times of events are recorded to full computer accuracy, rather than simply as taking place in a given year;
- DES allows realistic modelling of the time taken to report screening results and the possibility that this could be influenced by the number of screenings performed in a given time period.

The AHCPR model contains 20 states, of which one can only be reached by screening. The remaining 19 states may conveniently be divided into pre-invasive and invasive states.

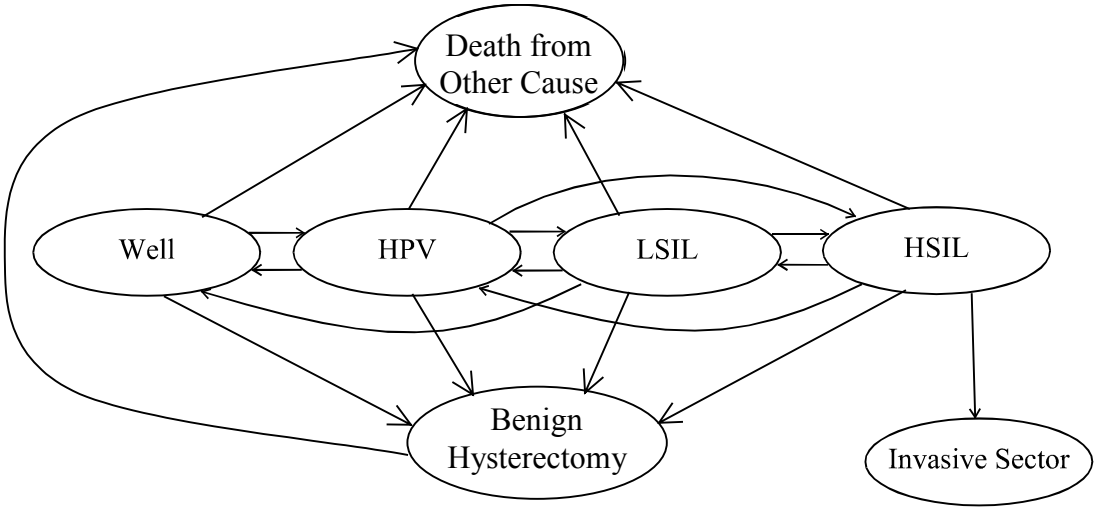


Figure 1 The pre-invasive sector of the natural history model

The possible state transitions in the pre-invasive sector are shown in Figure 1. Women may progress from the state "Well" through acquisition of human papillomavirus ("HPV") to low-grade squamous intraepithelial lesion ("LSIL") and high-grade squamous intraepithelial lesion ("HSIL"). We have maintained the assumption from the AHCPR model that HPV is a necessary precursor of SIL. The transitions shown in Figure 1 are those included in the AHCPR model. However, because of the different way of handling time in our model as described above, it is possible to progress from Well to HSIL (or even to Invasive) within one year. To maintain a measure of compatibility with the AHCPR model, we have included direct transitions between HPV and HSIL, and from LSIL to Well.

In addition, women may undergo hysterectomy for cause other than cervical cancer or SIL; such events are referred to an benign hysterectomies. Finally, all women are followed through to death. Death from "other causes" is possible in any state.

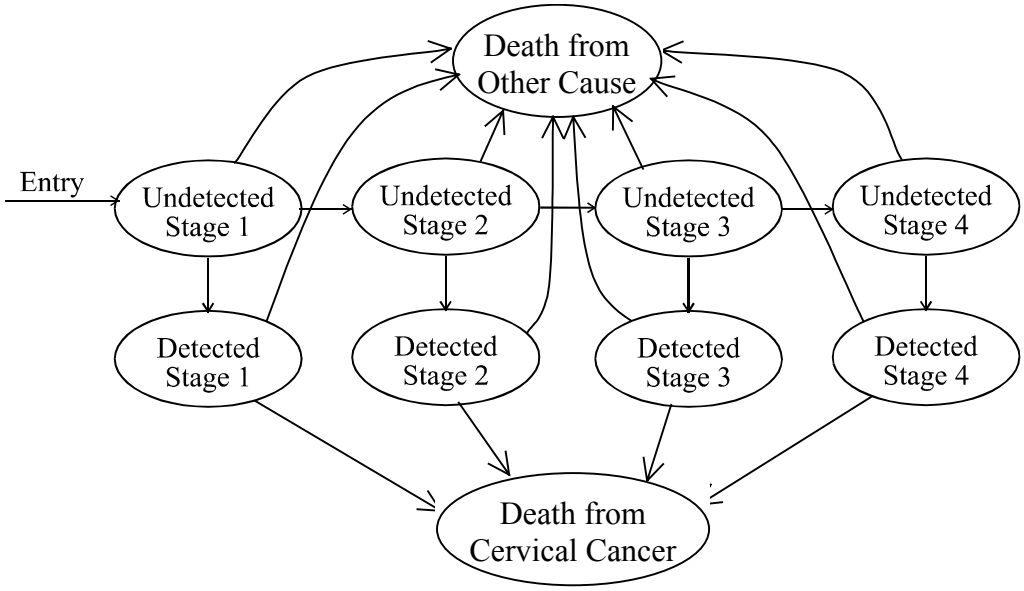


Figure 2 The invasive sector of the natural history model

Figure 2 shows the transitions in the invasive sector. Entry to this sector is by progression from HSIL to undetected cervical cancer. Cervical cancer is taken to have four stages. As long as the cancer remains undetected, progression to further stages is possible. The cancer may be detected either through screening or clinical signs. Once the cancer has been detected, it is assumed that treatment starts, and that no further progression is possible. In our model, the only event following detection of cancer is death, which may be either from cervical cancer or from other causes. The AHCPR model considers all deaths more than 5 years after detection of cancer to be deaths from other causes and includes "cancer survivor" states for women more than 5 years after detection.

An important assumption in this part of the model is that all cancers become symptomatic before death, and are thus detected and treated. The structure in our model could allow for a difference in life expectancy according to time spent in the undetected stage, and according to whether detection was through screening or symptoms. In the absence of the necessary data, this has not been done in the version of the model presented here. It would also be possible to allow for death from cervical cancer to follow directly from an undetected cancer; again, this has not been done in this model.

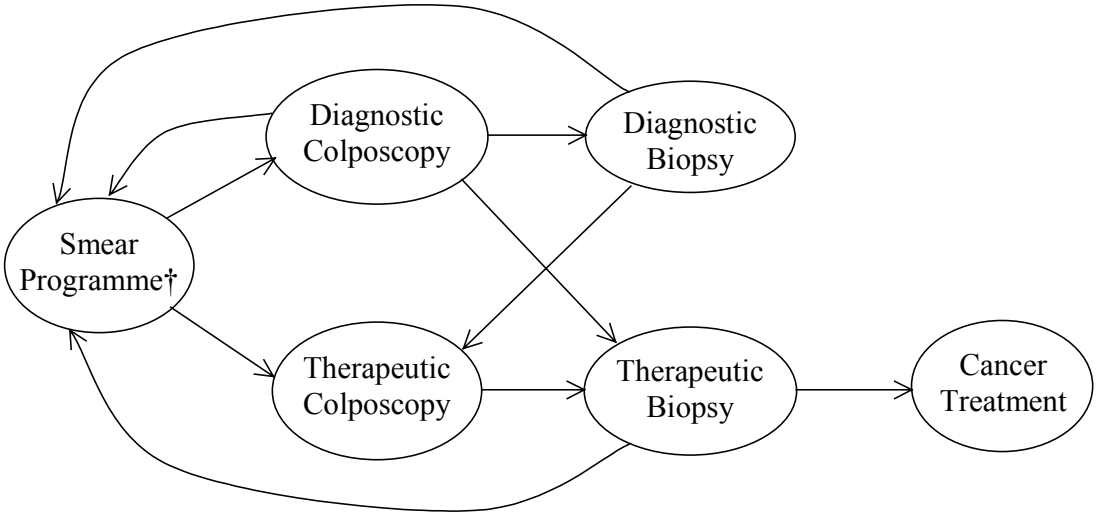
Screening procedure - outline

Any individual woman will enter the screening programme at age 20 with her first smear test. After any smear test, she will be waiting for the result. As long as the smears return a normal result, she will continue the routine cycle of a smear at regular intervals (in the base case analysis of the model, every 5 years). If the smear is inadequate, or shows borderline or low-grade dyskaryosis, she will be under surveillance in the screening programme: she will then have a repeat smear after a shorter recall time. This surveillance continues until either she has a normal smear result, in which case she returns to routine screening, or sufficiently many non-normal results (details appear later in this section), in which case she will be referred for colposcopy, possibly leading to biopsy and treatment. A single smear showing high-grade dyskaryosis is sufficient for referral to colposcopy. After colposcopy and possible biopsy, a woman may return to the smear programme under surveillance, until a sufficient number of normal smears have occurred, when she returns to routine. The woman remains in the screening programme until one of the following occurs:

- Death from other cause
- Benign hysterectomy
- Detection of cervical cancer
- Reaching the age of at least 65 in routine smearing

Note that a woman reaching the age of 65 under surveillance remains in the screening programme as long as surveillance is thought necessary.

From a modelling point of view, the important distinction is whether the next screening event for the woman is a smear test or result, or colposocopy or biopsy. This is illustrated in Figure 3



† This includes both routine and surveillance recall

Figure 3 Stages of the screening programme

Details of the screening programme

Each time a smear is taken, the result is recorded on a scale normal, inadequate, borderline, low grade or high grade. A high grade smear leads immediately to referral for further investigation. Other results contribute to a "cytological score": when this score reaches an appropriate level, referral for further investigation is made.

Sasieni *et al*[10] have used a simple score of no points for a normal smear, 1 point for borderline, 2 for low grade and 3 for high grade, with a referral threshold at 3 points. Two successive normal smears cancel out the existing score. In practice, women may be referred, even after a normal smear, for reasons other than the smear result: we have excluded such referrals from the model.

This system does not allow for referral after repeated inadequate smears, as required by the internally derived patient walkthrough. Our points system thus allows for referral after three inadequate or borderline smears, two low grade or one high grade: in our system, a borderline smear followed by a low grade smear does not lead to referral for women with no previous diagnosis of CIN. To take account of the above, we use the scores in Table 1.

Smear result	Cytological score
Normal	0
Inadequate	2
Borderline	2
Low-grade	3
High-grade	6

Table 1 Cytological scores used for deciding referral in model

Women returned to the smear programme after colposcopy or biopsy are under surveillance and will be referred back to colposcopy after one low-grade or two inadequate or borderline smears. This is modelled by setting the cytological score to 3 points before the first smear taken under surveillance.

The timings of the screenings are based on six parameters as shown in Table 2. For ease of reading, the default settings are used in the following description, although it is emphasised that they can all be changed.

Parameter	Default setting
Earliest screening age	20 years
Latest screening age	65 years
Routine recall time between screenings	5 years
Possible recall times between screenings while under surveillance	1 year
	6 months
	3 months

Table 2 Screening times

Provision is made for a limited capacity at the laboratory where smear reading takes place. This is done by having a minimum time between issue of successive smear results. The minimum time is one year divided by the annual capacity. When a smear is taken, the time for the result is set to be either 2 weeks after the smear was taken, or the minimum interval after the previous scheduled result time, whichever is the later. If a substantial queue builds up, it is possible for the woman to change state before the

smear is read. The result depends on her state at the time the smear was taken, not at the time it is read.

The recall times between screenings are taken from the most recent smear or colposcopy; if the queue has become so long that the next smear is overdue at the time of issue of a smear result, the next smear is taken immediately. If a smear result leads to referral to colposcopy, this is assumed to take place 2 weeks after issue of the smear result; we are assuming no delays due to shortage of capacity anywhere other than smear reading.

Individual characteristics

At any time, an individual woman in the screening programme has a cytological score and a recall programme. Initially, the cytological score is set at zero, and the recall programme is set to "every 5 years". The first smear takes place at age 20.

Smear testing

The result may be any of the following:

- *Normal* - set next smear according to the individual woman's recall programme, unless that would take the woman over the age of 65 and in routine recall, in which case no further screening takes place. If this is the second successive normal smear, then the cytological score is set to zero.
- *Inadequate* – add 2 to the cytological score, and set next smear for 3 months' time, unless the cytological score is at least 6, in which case refer for diagnostic colposcopy.
- *Borderline* - add 2 to the cytological score, and set next smear for 6 months' time, unless the cytological score is at least 6, in which case refer for diagnostic colposcopy.
- *Low-grade* – add 3 to the cytological score, and set next smear for 6 months' time, unless the cytological score is at least 6, or there is a previous diagnosis of CIN at biopsy, in which case refer for diagnostic colposcopy.
- *High-grade* - refer immediately for therapeutic colposcopy.

Diagnostic colposcopy

Visual inspection may reveal the cervix to appear either normal or atypical.

- *Normal* – set cytological score to 3. Immediate smear is taken and referred for result as above. Two further recalls at 6 month intervals before return to routine.
- *Atypical* – immediate therapeutic biopsy if previous diagnosis of CIN at biopsy, diagnostic biopsy otherwise.

Diagnostic biopsy

It is assumed that biopsy is 100% accurate diagnostically and has no effect on the actual condition of the woman, either immediately or in terms of future progression. Action taken is as follows:

- *Well or HPV* – set cytological score to 3, individual recall programme to one recall at 6 months, two recalls at 1 year, then return to routine. Next smear at 6 months according to this programme.
- *LSIL* – record diagnosis of CIN at biopsy, set cytological score to 3, individual recall programme to one recall at 6 months, two recalls at 1 year, then return to routine. Next smear at 6 months according to this programme.
- *HSIL or Invasive* – therapeutic colposcopy.

Therapeutic colposcopy

This is the first part of a process leading to therapeutic biopsy: it is recorded as a separate stage so that diagnostic colposcopy can lead directly to therapeutic biopsy as described above.

Therapeutic biopsy

It is assumed that biopsy is 100% accurate diagnostically. If it shows cancer, then the woman is moved from undetected to detected stage *n* and treatment for cervical cancer follows.

Otherwise, the action taken is as follows, depending on the state of the woman before treatment:

- *Well or HPV* – set cytological score to 3, individual recall programme to one recall at 6 months, two recalls at 1 year, then return to routine. Next smear at 6 months according to this programme.
- *LSIL* – record diagnosis of CIN at biopsy, set cytological score to 3, individual recall programme to one recall at 6 months, two recalls at 1 year, then return to routine. Next smear at 6 months according to this programme. Following the treatment, the woman may be in state Well, HPV or LSIL.
- *HSIL* – record diagnosis of CIN at biopsy, set cytological score to 3, individual recall programme to two recalls at 6 months, four recalls at 1 year, then return to routine. Next smear at 6 months according to this programme. Following the treatment, the woman may be in state Well, HPV or HSIL.

Parameters used in the model

Parameters in the model were derived from the best available sources in each case. The parameters used, together with their sources, are listed in the forthcoming HTA report. Since the purpose of this paper is to discuss the potential future use of the model, and for reasons of space, the actual values of the parameters used have been omitted.

Sample results

The model was run for a number of populations with an arrival rate of 1000 new women per year. The screening capacity was set to the smallest multiple of 1000 screens per year which would allow the model to run without the queue becoming unacceptably long. In all the runs reported here, we used a capacity of 11000 per year.

Results were obtained that could feed into a cost-effectiveness analysis. On the cost side, the number of each procedure carried out in each year of simulated time was collected. For effectiveness, in the absence of any data on quality of life we worked simply in terms of life-years saved. In this model, the age at which each woman would have died from other causes was known. Each time a death from cervical cancer occurred, the years of life lost could be determined.

Because the model is a stochastic model, the results will be slightly different each time that different sets of random numbers are used. Also the model starts empty so that it is necessary to run for a "warm-up" period until the model reaches a steady state. The figures reported in each table are averages per year over three runs each of 100 (simulated) years running of the model after discarding a warm-up period of 100 years. Quasi-standard errors for the difference are shown. These could be reduced simply by increasing the number of runs of the model.

For a base case run, we have assumed that the sensitivity of AutoPAP is 10 per cent higher than conventional testing, conditional on an adequate smear. The results are shown in Table 3.

	No AutoPAP	With AutoPAP	Difference	QSE(diff)*
Screenings	10009.8	10063.0	53.2	8.3
Normal/inadequate	9722.2	9758.8	36.6	7.5
Borderline/low/high	287.6	304.2	16.7	1.5
Colposcopies	157.6	170.1	12.5	1.1
Diagnostic Biopsy	11.4	13.3	1.9	0.3
Therapeutic Biopsy	104.7	114.1	9.4	0.9
Cancer treatments:				
Stage 1a1	1.987	1.670	-0.317	0.106
Stages 1a2 and 1b1	1.620	1.413	-0.207	0.106
Stages 1b2, 2, 3, and 4	3.560	2.693	-0.867	0.150
Life years lost**	215.2	168.4	-46.8	7.5

NOTES:

* QSE = quasi standard error

** Each time a death from cervical cancer was recorded in the model, the number of life years lost for that woman could be found by subtracting her age at death from the age at which she would have died from other causes. This approach has the advantage over measuring average age at death that variation in normal life-time is eliminated from the variance of the estimate, thereby increasing the precision of the result from a small number of runs of the model.

Table 3 Sample base case results from the model

As an example of a sensitivity analysis, the sensitivity and specificity for diagnostic colposcopy were both reduced from 100 per cent to 80 per cent. The differences were within the range of random error. This is not surprising, as the number of diagnostic colposcopies is quite small.

Validation

The validity of a model may be tested by comparing the output from the model with data about the "real" system being modelled. One comparison that can be made is the age-related incidence of cervical cancer. Yearly rates for incidence and detection of cancer are plotted in Figure 4 along with English cancer registry data for 1998[9], which

are provided in five-year age bands. As can be seen, detection has sharp peaks at the standard screening ages. A better comparison can be seen in Figure 5, where the detection rate in the model is adjusted to five-year bands.

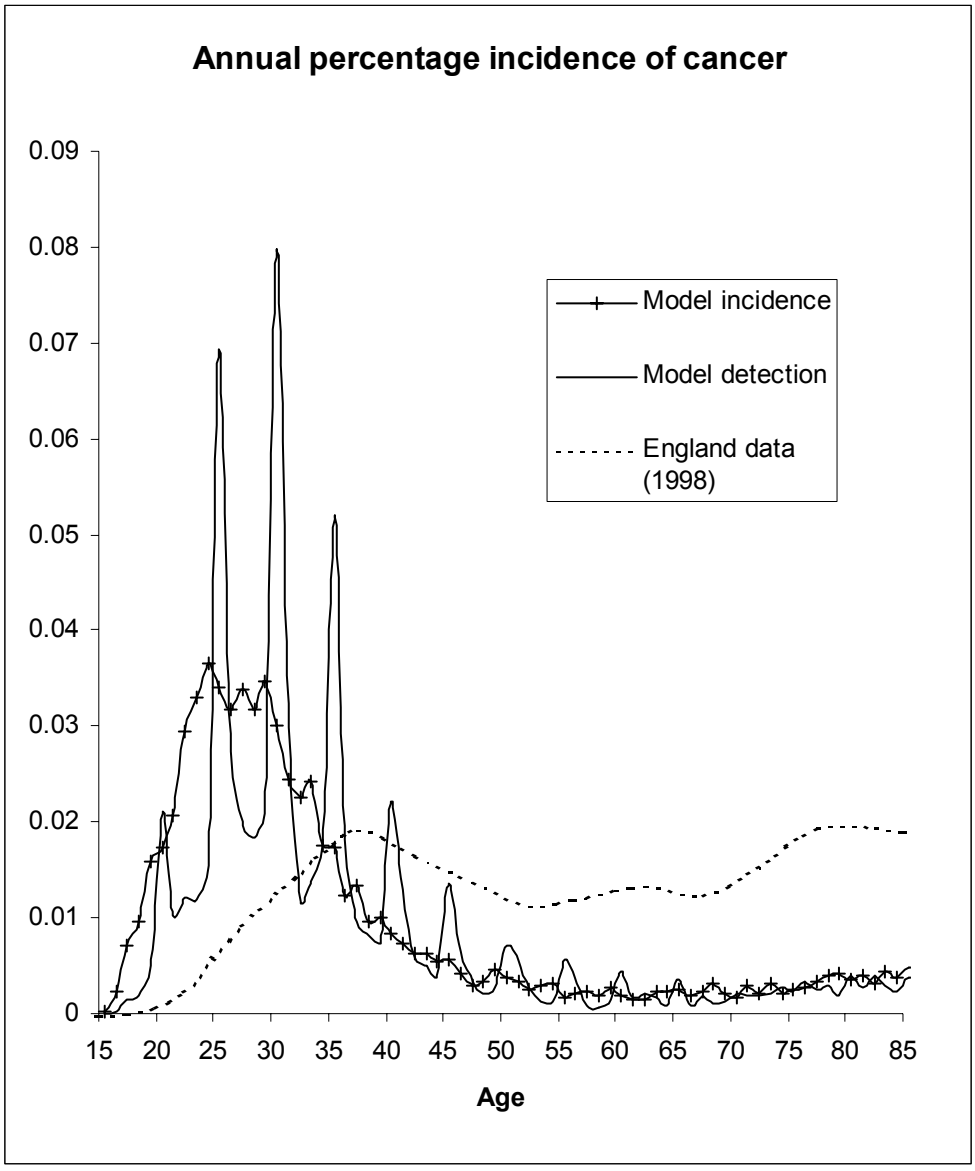


Figure 4 Age-related incidence and detection of cancer

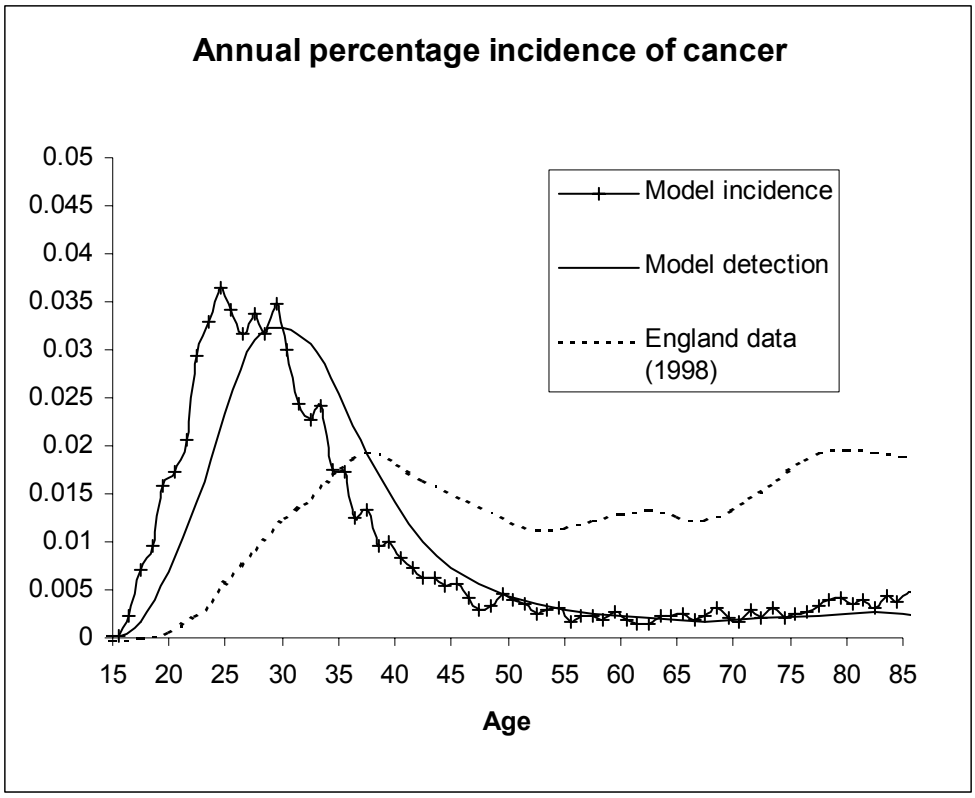


Figure 5 Age-related incidence and detection of cancer (model rates by 5 year bands)

Comparing the model output with the English data, it is clear that cancers are occurring in the model at a somewhat earlier age than in reality. However, the comparison at least matches the general shape in that there is a peak age of incidence. It is possible that changes in sexual behaviour mean that the model is a better predictor of the future age of peak incidence than the current data. Comparing with other UK models, Payne *et al*[7] had a steadily increasing incidence throughout the screening age range, and Cuzick *et al*[8] did not give a comparison, stating: "Validation studies were not in the scope of this review project."

In summary, we remain concerned about the validation of the model. We suspect that the problems experienced relate to being compelled to use data for particular parameters that may not be entirely appropriate. For instance, data on HPV incidence and progression tend to be derived from higher risk populations which may not be typical of the screened general population whose cancer incidence and smear results we are attempting to replicate.

Capability of the model

This model is an improvement on previous models in that it does not impose specific shapes on the survival times in any given state. It also allows for inadequate smears and takes realistic account of the timing of repeat smears. It is thus able to accommodate guidelines for repeat and surveillance smearing.

As it stands, the model is largely using data collected for models which have more restrictive assumptions placed upon them. In principle, the model can allow changes in a woman's condition to be dependent on previous history in any way that can be specified. In particular, annual rates of progression can depend on any combination of the woman's age, how long she has been in her current state and how long she was in any previous state.

One particular matter that would be usefully addressed is the issue of survival times from detection of cancer. As the model stands, a woman with undetected stage 4 cancer will have her life expectancy reduced by earlier detection of the cancer, since survival time is effectively measured from detection. This only has a minor effect on the model. Earlier stages of cancer are also affected by this problem, but it is offset by the possibility that detection through screening could prevent progression to later stages.

Other possibilities for future development of the model include assessment of the transient effects of a change in screening policy. The model can be set to reach steady state under one screening policy and then introduce a new policy from a given calendar date. The short term effects and time taken to reach a steady state can be shown. The model could also allow for a growing population by increasing the number of women entering the model each year. This would require some method for systematically increasing the screening capacity with the population if waiting times for screening results are not to increase without limit.

Where to go next?

We have developed a simulation model whose structure is capable of answering a wide range of questions concerning cervical screening. The specific issue for which it was developed was a comparison between partly automated and fully manual reading of smears taken in the same way. Our systematic review suggested that automation makes no difference to the sensitivity and specificity of the screening test. If that is the case, then the question of whether automation is advantageous can be settled by a cost-minimisation analysis. However, any sensitivity analysis to test the possibility that automation is better (or worse) than manual reading requires a full assessment of the implications of the change in test characteristics. This, in turn, requires a full simulation model.

To explore the full impact on costs and effectiveness associated with the introduction of automated technologies requires a model which adequately reflects the realities of the cervical screening programme. We have constructed a model which could be developed for this purpose, given adequate data. Such a model would also be able to address other questions with regard to the cervical screening programme, and we would thus suggest that its development should represent a high priority for further research.

References

[1] Secretary of State for Health. *Saving lives: Our healthier nation.* London: HMSO, 1999

[2] Office for National Statistics. Mortality statistics; cause, 1997. Series DH2, number 24. London: The Stationery Office, 1999.

[3] Fidler HK, Boyes DA, Worth AJ. Cervical cancer detection in British Columbia. A progress report. *Journal of Obstetrics and Gynaecology in the British Commonwealth* 1968;**75**:392-404.

[4] Van Oortmarssen GJ, Habbema JD. Epidemiological evidence for age-dependent regression of pre-invasive cervical cancer. *British Journal of Cancer* 1991;**63**:559-665.

[5] McCrory DC, Mather DB, Bastian L, Datta S, Hasselblad V, Hickey J *et al*. *Evaluation of Cervical Cytology.* Evidence Report/Technology Assessment No 5. AHCPR Publication No 99-E010. Rockville, MD: Agency for Health Care Policy and Research, February 1999.

[6] Report of a Working Party set up by RCPath, BSCC and NHSCSP. *Achievable standards, benchmarks for reporting & criteria for evaluating cervical cytopathology.* Sheffield: NHSCSP Publication No 1, 2nd Edition, May 2000.

[7] Payne N, Chilcott J, McGoogan E. Liquid-based cytology in cervical screening: a rapid and systematic review. *Health Technology Assessment* 2000;**4**(18).

[8] Cuzick J, Sasieni P, Davies P, Adams J, Normand C, Frater A *et al*. A systematic review of the role of human papillomavirus testing within a cervical screening programme. *Health Technology Assessment* 1999;**3**(14).

[9] Office for National Statistics. Cancer statistics; registrations. Registrations of cancer diagnosed in 1998, England. Series MB1, no.29. London: Office for National Statistics, 2000.

[10] Sasieni P, Cuzick J, Lynch-Farmery E and the National Co-ordinating Network for Cervical Screening Working Group. Estimating the efficacy of screening by auditing smear histories of women with and without cervical cancer. *British Journal of Cancer* 1996;**73**:1001-1005.