Monitoring and evaluating the performance of the UK NHS Cervical Screening Programme: monitoring performance by using cytology outcomes adjusted for population characteristics

R. G. Blanks, M. Waller, A. Sanchez-Galvez and S. M. Moss

Cancer Screening Evaluation Unit, Institute of Cancer Research, Sutton, Surrey, UK

Accepted for publication 16 December 2003

R. G. Blanks, M. Waller, A. Sanchez-Galvez and S. M. Moss

Monitoring and evaluating the performance of the UK NHS Cervical Screening Programme: monitoring performance by using cytology outcomes adjusted for population characteristics

Current quality assurance measures used in the NHS cervical screening programme (NHSCSP) include a review of laboratories with percentages of moderate/severe and borderline/mild smear results outside the 10th–90th percentiles. The method is limited by the fact that many of these outlier smear percentages may reflect laboratories covering populations with low or high risk and/or short or long average screening intervals. This paper outlines a new approach to aid the detection of outlier laboratories, by using data collected at the primary care trust (PCT) or health authority (HA) level and making allowances for population characteristics and screening interval. The setting is the NHSCSP in England using annual data provided by HAs. Data from the screening year 2000–01 is used to illustrate the methodology, although the methods can also be applied to data at the PCT level (now being collected for 2002–03 onwards). Percentages of smear results have been analysed against a series of explanatory variables using logistic regression models. These explanatory variables include Townsend deprivation index, uptake-corrected ethnic minority composition, a measure of screening interval, area type and region. An expected percentage of borderline/mild and moderate/severe smears is estimated from the models and an observed : predicted ratio ($OPR^{mod/sev}$ and $OPR^{bord/mild}$) calculated. Low values are suggestive of relative undercalling and high values overcalling, after allowance for population characteristics. Analysis of data for 2000–01 showed that the $OPR^{mod/sev}$ for the 99 HAs varied from 0.68 to 1.44. Laboratories with low percentages of moderate/severe smears, but associated with PCTs or HAs with $OPR^{mod/sev}$ values closer to unity may not need to be investigated as their observed rates are consistent with predicted rates based on population characteristics. The method could also be directly applied to laboratories if further information on the population covered by each laboratory were routinely collected.

Keywords: cervical screening, performance, cytology, population characteristics, deprivation, ethic minority composition

Introduction

The evaluation of cervical screening is a complex task, particularly because of the absence of any randomized controlled trials with which to compare performance outcome measures. Epidemiological methods currently being used to evaluate the performance of the UK NHS cervical screening programme (NHSCSP) include using cases of invasive cervical cancer and comparing their screening histories with age and

Correspondence:
R. G. Blanks, Cancer Screening Evaluation Unit, Brookes-Lawley Building, Institute of Cancer Research, Cotswold Road, Sutton, Surrey SM2 5NG, UK.
Tel.: +44 (0) 20 8722 4008; Fax: +44 (0) 20 8770 0802;
E-mail: r.blanks@icr.ac.uk

residency-based controls to evaluate the efficacy of screening.[1] Whilst such studies form the most definitive evidence of a national screening programme's performance, at the local level they have two limitations. First, disease incidence reflects performance some time in the past, and secondly, the numbers of cases in any one Health Authority (HA) in a given year will be small, leading to problems of statistical instability. Such studies are therefore less useful for producing rapid assessments of local screening performance. Other methods are therefore required to provide evidence on performance at the local level, which can be used to trigger further evaluation of performance in a proactive and timely manner.

This paper will use the information on cytological smear outcome (grouped as borderline/mild dyskaryosis or moderate/severe dyskaryosis) for women resident in the 99 HAs in England in the screening year (1 April–31 March) 2000–01 to calculate observed and expected percentages of smear results. The basic information is returned annually by HAs (KC53 return) to the Department of Health and information is published annually in the Department of Health Statistical Bulletin.[2] From 2002–03 KC53 returns will be compiled by PCT rather than HA, although the methods discussed in this paper apply equally. In general HA returns include test results from more than one laboratory. Information from laboratories, which is based on tests (rather than women as for the KC53) are returned using the KC61 form. The percentage abnormal smears from KC61 returns relate to GP/community clinic tests but exclude hospital cases.

Current laboratory performance statistics targets for borderline/mild and moderate/severe smear percentages are based on calculating the 10th and 90th percentiles of the distribution from all laboratories in England. Laboratories with performance outside this range are required, with the assistance of their quality assurance team, to investigate and support the explanation for this performance. Table 1 shows the 10th–90th percentile figures for data from 2000–01, and the use of these criteria are discussed further in the NHSCSP document on achievable standards.[3] However as the background risk of cervical cancer incidence varies by a factor of three,[4] high or low percentages of moderate/severe smears may simply represent laboratories covering populations with very low or high risks. Furthermore the percentage of borderline/mild or moderate/severe smears expected will also depend on the screening interval, with a higher rate in women screened every 5 years than women screened every 3 years.

The method presented in this paper takes into account the population characteristics of women screened and screening interval using information that is presented at the HA/PCT level to calculate expected percentages. If the percentage of smears analysed by a laboratory coming from each HA/PCT were known, a weighted average of the expected percentages at the HA level could then be applied to the laboratory to give laboratory expected percentages. These data are at present not routinely collected and this paper will therefore primarily consider the results at the HA/PCT level and how this can be applied to indirectly assess laboratory performance.

Table 2 shows an extract of the data published in 2000–01 from five of the 99 HAs with the smear results expressed as a percentage of all adequate smears. The Table shows the HAs with the two lowest, the median and the two highest percentages of negative smears, anonymized as HAs A, B, C, D and E respectively. The percentage figures are based on large numbers (and are therefore statistically stable) and there is no substantive confounding by age, as the age distribution of screened women is similar across all HAs. The percentage of negative smears ranges

Table 1. Current laboratory target range (10th–90th percentile) for smear results from laboratories in screening year 2000–01

	10th–90th percentile
Mild/borderline as % of all adequate smears	4.2–9.4%
Moderate or worse as % of all adequate smears	0.9–1.8%

Table 2. Extract from Cervical Screening Programme Statistical Bulletin 2000–01 showing Health Authority percentages of negative, borderline/mild and positive smears ranked by increasing percentage of negative smears

Health Authority	Rank	% Negative	% Borderline/mild	% Moderate/severe
A	1	85.99	11.90	2.11
B	2	87.79	10.48	1.73
C	50	93.29	5.50	1.21
D	98	95.68	3.11	1.22
E	99	95.78	3.16	1.06

from about 86% in A HA to nearly 96% in E HA. The percentage of moderate/severe smears (which leads directly to colposcopy) is twice as high in A as in E and the percentage of borderline/mild smears three times higher. Women with repeat borderline/mild smears will also be referred to colposcopy, so these figures give some information about the colposcopy services required in the various HAs. However, without allowance for population characteristics and screening interval it is not clear, for example, whether the high moderate/severe percentages reported by HAs A and B are likely to be suggestive of overcalling or not.

Methods

Data on observed proportions of borderline/mild or moderate/severe smears from the 99 HAs are used as the outcome variable in a regression model with a number of explanatory variables. The observed proportions for each of the 99 HAs are then compared with the expected values from the model. The results can be expressed as an observed versus predicted ratio (OPR) for moderate/severe smears ($OPR^{mod/sev}$) and borderline/mild smears ($OPR^{bord/mild}$).

The models have been constructed by using multiple logistic regression with proportion of borderline/mild smears or proportion of moderate/severe smears as the response variable. The explanatory variables used in the models can be divided into three types, first population characteristics such as deprivation and ethnicity, secondly screening interval, and thirdly characteristics such as region and area type (city, urban or rural),

Region and area type are categorical explanatory variables and deprivation index, ethnic correction factor and screening interval measure (SIM) are entered as continuous explanatory variables. The regression models for borderline/mild and moderate/severe have been run separately to avoid constraining the relation between borderline/mild and moderate/severe results.

Description of explanatory variables

Region. Each of the 99 HAs is from one of eight regions, Eastern, London, North West, Northern & Yorkshire, South East, South West, Trent and West Midlands, which were current in 2000–01.

Area type. One of mainly city, mainly urban and mainly rural. Of the 99 HAs 35 (35.4%) have been classified as mainly rural, 51(51.5%) as mainly urban and 13 (13.1%) as mainly city.

Townsend deprivation index. A mean deprivation index has been produced for the HAs using data originally classified at the PCT level available from the National Database for Primary Care Groups and Trusts. The deprivation index is a continuous variable ranging from −3.7 the least deprived, to +11.9, the most deprived, areas with high levels of deprivation being associated with a higher risk for cervical cancer.[5]

Ethnicity. Different ethnic groups are at differing risk for cervical cancer; in particular the Black Caribbean population are at high risk whilst the Indian subcontinent population are at low risk.[6] These groups will also have different probabilities of attending screening and the populations have therefore been adjusted for relative uptake before being entered in the model. The adjustment for attendance probability is particularly important for the Pakistani and Bangladeshi groups in the Indian subcontinent population because of the considerably lower published attendance rates from these groups.[7] The percentages are entered on the log scale to normalize the log-normal (positively skewed) distribution of the percentage of ethnic minority composition for the 99 HAs. For the majority of HAs the ethnic minority composition is less than a few percent. The population of women in each ethnic group is taken from the National Database for Primary Care Groups and Trusts. The method is discussed further in the appendix.

Screening interval measure. This is derived from the KC53 return supplied by HAs and is an estimate of screening interval calculated using the number of women screened in the previous 3 years divided by the number of women screened in the previous 5 years, termed PROP3–5 or P. This will only give an approximate measure of mean screening interval as peak invitations to screening can occur at 3 or 5 years from the previous negative screen, within a single HA and an additional underlying continuous distribution of intervals is present which result from 'opportunistic' smears (these are smears initiated by the GP, or the woman, and not by the call/recall system). The SIM can be estimated as $8–5 \times P$. At one extreme if all women had been screened within the last 3 years, then $P = 1$ and $SIM = 8-(5 \times 1) = 3$. In practice an HA with a 3-year screening interval would not be expected to have an SIM of exactly 3 as recall is

Table 3. Demographic and population characteristics and screening interval of example Health Authorities

Health Authority	Region	Area type	Deprivation index	Screening interval measure	Percent Black Carib*	Percent Ind Sub*
A	North West	Urban	+3.55	4.26	0.06	0.13
B	London	City	+8.68	4.06	3.68	2.43
C	Eastern	Rural	−1.89	4.47	0.05	0.13
D	Northern & Yorkshire	Rural	+0.98	3.67	0.04	0.14
E	Northern & Yorkshire	Urban	+2.86	4.45	0.73	8.14

*Adjusted for attendance probability.

carried out at 3 years rather than a few months earlier. At the other extreme for a true 5-year interval, 60% of women will have been screened in the previous 3 years, therefore $P = 0.6$, and $SIM = 8-(5 \times 0.6) = 5$. In practice in 2000–01 SIM varied from 4.82 years to 3.56 years. The data suggest that those HAs with a 5-year policy have in most cases an interval that is much less than this and further analysis shows this to be the result of a greater usage of 'opportunistic' smears. For 2000–01 results suggest that one HA with a 5-year policy (but a substantial usage of opportunistic smears) actually had an SIM less than that of one HA with a 3-year policy. The screening interval policy of a HA is therefore not necessarily a good estimate of the screening interval used in practice because of the substantial use of opportunistic smears in many HAs, particularly those with a 5-year screening policy.

To illustrate these explanatory variables further, Table 3 shows the same HAs as Table 2 with information about the explanatory variables. A and B, the two HAs with the lowest percentages of negative smears, have populations with risk factors, particularly deprivation index, suggesting a higher risk of cervical cancer.

Results

The regression models for borderline/mild and moderate/severe smears are shown in Table 4. The SIM and deprivation index are both highly significant variables whereas area type (city/urban/rural) has less effect after allowing for the other variables. Ethnicity is a significant predictor of borderline/mild smears but less important for moderate/severe smears with data at the HA level. Ethnicity is likely to be more important when data are analysed at the smaller PCT level. About 50% of the variation in smear proportions are explained by these variables. The other 50% will be due to relative levels of laboratory overcalling or undercalling and the inadequacy of the explanatory variables in explaining all the risk for non-negative smears. The models show as expected that the risk of obtaining a borderline/mild or moderate/severe result increases with screening interval [an odds ratio (relative risk) of 1.21 per year for moderate/severe smears]. There is a relative risk of 1.03 per unit increase in deprivation from −3.7 to +11.9. The model also shows that the North West region contains the population with the highest risk of a moderate/severe test result.

Table 5 shows the observed and predicted numbers and percentages for the five illustrative HAs. Both B

Table 4. Multiple logistic regression analysis of borderline/mild and moderate/severe proportions of smears against explanatory variables

	Borderline/mild		Moderate/severe	
Variable	Odds ratio	*P*-value	Odds ratio	*P*-value
SIM	1.15	<0.001	1.21	<0.001
London	1.12	<0.001	1.00	0.932
North West	1.17	<0.001	1.28	<0.001
North & Yorkshire	0.85	<0.001	1.00	0.862
South East	0.96	<0.001	1.05	0.029
South West	1.04	0.004	1.10	<0.001
Trent	0.86	<0.001	1.10	<0.001
West Midlands	1.03	0.034	0.98	0.318
Urban	1.00	0.534	1.02	0.212
City	1.12	<0.001	1.04	0.165
Townsend	1.03	<0.001	1.03	<0.001
*Ethnic: log BC%	1.02	<0.001	1.01	0.289
*Ethnic: log IS%	0.96	<0.001	0.99	0.053

Baseline for categorical variables are 'Eastern' for Region and 'Rural' for area type (odds ratio = 1).

*BC, uptake adjusted Black Caribbean population; IS, uptake adjusted Indian subcontinent population.

Table 5. Observed and predicted numbers and percentages of borderline/mild and moderate/severe smears from example Health Authorities expressed as a ratio [observed/predicted ratio (OPR)]

	Borderline/mild numbers (%)			Moderate/severe numbers (%)		
Health Authority	Obs (%)	Pred (%)	OPR	Obs (%)	Pred (%)	OPR
A	2441 (11.9)	1571 (7.7)	1.55	433 (2.1)	408 (2.0)	1.06
B	3043 (10.5)	2619 (9.0)	1.16	501 (1.7)	512 (1.8)	0.98
C	2026 (5.5)	2120 (5.8)	0.96	447 (1.2)	505 (1.4)	0.89
D	1219 (3.1)	1882 (4.8)	0.65	478 (1.2)	498 (1.3)	0.96
E	744 (3.2)	1199 (5.1)	0.62	249 (1.1)	359 (1.5)	0.69

and C HA have predicted and observed values that are very similar, suggesting no evidence of overcalling or undercalling relative to the average values expected from HAs with similar population characteristics. Whilst B HA has one of the highest percentages of borderline/mild and moderate/severe smears of all 99 HAs it also has one of the highest expected percentages. There is therefore no evidence of overcalling after allowing for population characteristics and screening interval. In contrast A HA shows evidence of relative overcalling even after these allowances. This overcalling can be seen to relate almost exclusively to borderline/mild smears with 11.9% observed compared with only 7.7% predicted by the model, an $OPR^{bord/mild}$ of 1.55. The conclusion for A HA is that for 2000–01 there is some strong evidence that relative to all other HAs the associated laboratories may be overcautious in interpreting smears as borderline/mild rather than negative. The relevant laboratory details could be examined to determine the cytology/histology correlation for women with repeat borderline/mild results leading to colposcopy. If a very high proportion of these smears have negative histological findings then this would be further evidence of overcalling.

At the other end of the spectrum are D HA and E HA, which are the two HAs with the lowest percentages of non-negative smears. Both have lower percentages than predicted of borderline/mild smears and E HA has a lower percentage of moderate/severe smears.

With an $OPR^{mod/sev}$ of 0.62 and an $OPR^{bord/mild}$ of 0.69, E HA shows the strongest suggestive evidence of undercalling moderate/severe smears whereas the other four HAs have percentages of moderate/severe smears broadly in keeping with the modelled expected values. As undercalling of moderate/severe smears could result in a lower detection of CIN III, the data suggest that E HA has the strongest evidence for further investigation of the five HAs shown.

When considering all 99 HAs the OPR range for moderate/severe smears for 2000–01 was 0.68–1.44. The level of statistical stability of the OPR values for almost all HAs is reasonably high even with only 1 year's data. However, there are a few HAs with small populations where the data could more helpfully be combined over 2 years. The presentation of the data as observed and expected numbers allows the data to be readily combined over two or more years and OPR values calculated.

Validation of model

Some estimation of the validity of the approach can be obtained by comparing the OPR results in the (relatively unusual) circumstance where two HAs are covered by the same laboratory. The OPR values measured from the HAs should be similar. Table 6 shows two laboratories (1 and 2) which each process smears from two HAs (W,X & Y,Z respectively). There is evidence of the expected agreement, with laboratory 1 being seen to undercall both borderline/mild smears and moderate/severe smears as expressed by the low OPR values from HAs A and B. Conversely laboratory 2 tends to overcall borderline/mild smears and marginally overcall moderate/severe smears.

Table 6. Examination of data where more than one HA is known to be covered by the same laboratory

HA	Lab	$OPR^{bord/mild}$	Mean	$OPR^{mod/sev}$	Mean
W	1	0.62	0.69	0.69	0.80
X	1	0.75		0.90	
Y	2	1.42	1.49	1.08	1.10
Z	2	1.55		1.12	

When the new returns are available giving information by PCT there will be many PCTs where only one laboratory processes the smears and these comparisons will therefore be possible for many more laboratories. If we assume that the error in the population risk factors is random then the best estimate of the laboratory OPR is given by the average value of the estimates.

Discussion

This paper has demonstrated how a relatively simple model can be used to estimate the cytology outcomes from the local population characteristics and screening interval using annual data. Laboratories with very high or low reporting rates can then be examined further by seeing if these results are in keeping with that obtained at the HA or PCT level. If a laboratory with a low percentage of moderate/severe smears serves an area where there is a low predicted percentage of these smears then it can be argued that no further investigation may be required. In contrast if the local HA results also suggest undercalling, and hence a risk of a lower detection of CIN III, then this will strengthen the evidence for further investigation of the laboratory. In the context of reducing cervical cancer incidence and hence mortality, evidence of severe undercalling may be of more immediate concern than evidence of overcalling, although both should require further investigation.

It could be argued that it would be easier to use the background incidence of cervical cancer before the introduction of the call/recall system in 1988 (such as that estimated by Swerdlow and dos Santos Silva)[4] in place of demographic and population characteristics. The difficulty is that some screening has been in place in some parts of the country since the 1960s and therefore all useful measures of background incidence will be distorted by the presence of the screening programme itself. Furthermore as population characteristics change over time pre-screening background incidence would become increasingly inaccurate.

The current model used in this paper has three main highly significant explanatory variables (deprivation, ethnicity and SIM), which are risk factors for the outcome smear percentages. The other two variables may be additional surrogate measures for risk although the contribution to the model of area type is small. The inclusion of region in the model can be justified on the grounds of differing levels of risk related to different regional populations after allowing for deprivation and ethnic minority composition. These factors will be related to population or lifestyle characteristics between regions for each given level of deprivation and ethnicity. If region and area type are omitted from the model the conclusions for the majority of HAs are broadly the same, with the exception of one or two HAs in the North West region. It is hoped that improvements to the model can be produced in the future. In particular it may be possible to improve the precision of the ethnic minority correction factors. This may be aided by using information on the characteristics of the women actually attending screening rather than relying on the characteristics of the resident population. The model as it stands at present may however be reasonably adequate for its intended use.

Laboratories with, for example, low reporting levels of moderate/severe smears and associated with HAs/PCTs with low $OPR^{mod/sev}$ results could therefore be investigated to determine if there is further evidence of undercalling. Some of this information could be obtained from current laboratory QA schemes where a series of test smear readings are compared with consensus opinion.[8] Alternatively the population of women attending screening could be further investigated to determine if the characteristics of the screened population is different to that used in the model. Such an investigation could also help to improve the quality of information used in the model. It should be emphasized that the main purpose of the methodology is to enable a more proactive and rapid assessment of screening performance to be instigated enabling the targeting of smaller numbers of HAs for closer investigation. The final conclusion of any further investigations would either be that the associated laboratories to the HA do show evidence of relative over- or under-calling, or that the population characteristics used in the model do not fully reflect the attending women in that HA. This could be determined from questionnaire surveys of a random sample of attending women.

Ideally the OPR should be measured by laboratories to make it comparable with the current performance measures. To gain a measure of the OPR at the laboratory level for borderline/mild and moderate/severe smears it would be possible to use the expected percentage from HAs/PCTs and apply this to observed data on smear percentages directly from the laboratory returns (KC61s). However this requires an esti-

mate of the PCT(s) of origin of the smears analysed by each laboratory. This data is not currently routinely collected. However, the methodology can be used as presented in this paper, simply by knowing the approximate population area (PCTs) covered by each laboratory, and there are advantages to having the OPR calculation at both the HA/PCT level as well as the laboratory level.

In conclusion, these methods are aimed at further proactive quality assurance of the cervical screening programme. Any indications of the analysis represent suggestive evidence only and should be considered appropriately in that context. With the current pressures in the UK on individual-based epidemiological investigations, driven by confidentiality considerations, it is increasingly important that routinely available data are fully exploited.

References

1. Sasieni PD, Cuzik J. Lynch-Farmery 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. *Br J Cancer* 1996;**73**:1001–5.
2. Government Statistical Service. *Cervical Screening Programme, England*. London: Government Statistical Service; 2000–01. ISBN 1 84182 415 1.
3. NHS. *Achievable Standards, Benchmarks for Reporting and Criteria for Evaluating Cervical Cytopathology*. Johnson J, Patnick J (eds). London, UK: NHSCSP publication no. 1, May 2000.
4. Swerdlow A, dos Santos Silva I. *Atlas of Cancer Incidence in England and Wales 1968–85*. Oxford: Oxford University Press; 1993.
5. Harris V, Sandridge A, Black R, Brewster D, Gould A. *Cancer Registration Statistics: Scotland. 1986–1995*. Edinburgh: ISD Scotland Publications; 1998.
6. Bahl V. Cancer and ethnic minorities – the Department of Health's perspective. *Br J Cancer* 1996;**74**(Suppl. XXIX): S2–10.
7. Health Education Authority. *Black and Minority Ethnic Groups in England: Health and Lifestyles*. London: Health Education Authority; 1994.
8. NHS. *External Quality Assessment Scheme for Gynaecological Cytopathology*. London, UK: NHSCSP publication no. 15, January 2003.

Appendix

The relative risks and attendance probabilities taken from the literature[5,6] are shown in Table A.1. The uptake-adjusted percentage Black Caribbean population is entered in the model as one variable, and the Indian subcontinent population (Indian, Pakistani and Bangladeshi) as another.

Table A.1 Uptake-corrected ethnic minority variables

Ethnic minority	Relative risk	Relative attendance
White	1	1
Black Caribbean	1.12	1.05
Indian	0.66	0.82
Pakistani	0.66	0.58
Bangladeshi	0.66	0.43
Other	*	*

*No estimate made.

Table A.2 shows an example calculation from an HA with a particularly high ethnic minority composition. The population is multiplied by an overall attendance figure of 0.8 and the relevant relative attendance figure from Table 1 to get a population attendance estimate. The population attendance number for the Black Caribbean population is divided by the total population attendance to obtain the percent population attendance. Similarly, the population attendance numbers for the three Indian subcontinent (Indian, Pakistani and Bangladeshi) groups are summed and divided by the total population attendance to obtain the percent population attendance. The Log of the percent PA for Black Caribbean and Log of the percent PA for Indian subcontinent populations are entered into the model as two separate continuous variables.

Table A.2 Example calculations for a single Health Authority

Ethnic minority	Population	Attendance correction*	Pop-attend (PA)	% PA
White	26 5628	0.8	212502.4	
Black Caribbean	3 8334	0.84	32200.6	10.8
Indian	5237	0.66	3456.4	1.6
Pakistani	1614	0.46	742.4	
Bangladeshi	2057	0.34	699.4	
Other**	5 9439	0.8	47551.2	
Total			297152.4	

*0.8 × relative attendance.

**Other ethnic groups given a relative attendance of 1.