Does feedback improve the quality of cervical smears? A randomized controlled trial

FRANK BUNTINX

J ANDRÉ KNOTTNERUS

HARRY F J M CREBOLDER

TRUDIE SEEGERS

GERARD G M ESSED

HUBERT SCHOUTEN

SUMMARY. In a randomized controlled trial three methods of feedback of increasing intensity, directed at 183 doctors taking cervical smears, were compared with respect to their effects on the sampling quality of smears. Overall, feedback was found to have no influence on quality criteria in the crude data analysis. However, a significantly larger decrease in the percentage of smears lacking endocervical cells was found in the groups receiving monthly overviews of their results with peer comparison, when compared with the groups not receiving this type of feedback (odds ratio 0.75). Moreover, feedback appeared to have a clear effect on the presence of endocervical cells among doctors submitting a substantial number of smears in the intervention period, as opposed to those who submitted fewer smears. A positive correlation was also observed between the increase in the group mean of the proportions of smears containing pathological cells and the intensity of the feedback. However, this increase did not reach statistical significance.

This study suggests that monthly feedback with peer comparison may have a positive relationship with some aspects of quality improvement in cervical screening.

Keywords: cervical samplers; screening accuracy; cervical screening; diagnostic techniques; feedback.

Introduction

A RELATIONSHIP between indicators of the sampling quality of cervical smears and the probability of detecting pathology, both cytological and histological, has been reported by many authors.[1-6] Different opinions exist about the clinical significance of the presence of endocervical cells in the smear. It has been shown that absence of endocervical cells does not necessarily result in an excess of abnormal smears in the following years.[1] However, because of the demonstrated relationship between the presence of endocervical cells and the detection of pathology[1-6] the presence of endocervical cells is an important criterion for a good quality cervical smear.[7]

Improvement in the quality of cervical smears demands that those taking the samples should change their behaviour. Behavioural change presupposes that information is available, but information alone is not sufficient. In the field of cervical cytology, the provision of feedback to those taking samples could be expected to affect the quality of the smears submitted. Boon and colleagues have described an increase in the percentage of smears containing endocervical cells from 70% to 84% in the course of one year, following registration of the results of every sampler, and instruction by telephone of all those performing poorly.[8] Baker advised that reporting individual results should be supplemented with regular overviews allowing doctors to compare their results with those of colleagues.[9]

With the exception of these two reports, we have found no evaluation of the effect of feedback in this field. Therefore, three methods of feedback of increasing intensity have been compared in a randomized controlled trial to determine their effect on the quality of cervical smears obtained during routine consultations. This study was approved by the medical ethics committee of the Flemish Institute of General Practice.

Method

The study covered all smears obtained during routine consultations by doctors who had submitted at least 20 smears for cytological assessment during the year prior to the study (1988) to the department of pathology of the Central Laboratory of Antwerp, one of the larger cytological services in Flanders, Belgium (183 doctors out of a total of 238 submitting smears). Only those doctors submitting at least 20 smears were included in order to allow the observation of change in the quality of the smears subsequently submitted. The doctors included were not aware of the study during the study period.

For each doctor the following data were collected from various sources: sex, discipline (general practitioner or gynaecologist), province in which employed, and the number of smears submitted during 1988. At the start of the study all the doctors were telephoned and asked which device they used for obtaining a smear. The devices were classified into five groups: Cytobrush® (Medscand) only; Cytobrush® with a spatula; spatula only (one or two samples obtained with an Ayre-type spatula or with an extended tip spatula); two spatula samples obtained with the two ends of a combined spatula; and a miscellaneous group.

From 1 June 1989 until 30 April 1990 (the study period), all smears submitted to the department were subjected to a detailed quality assessment by one of four experienced cytotechnologists; smears were randomly allocated to the technologists. The cytotechnologists were aware of the study but did not know which form of feedback doctors had been assigned to receive.

Each smear was judged on a three-point scale (excellent, moderate or unsatisfactory) according to five criteria. A smear was considered 'adequate' if: endocervical cells, or squamous metaplasia, or both were present; the cells were so fixed that the contours of cellular structures were clear, subsequent staining was good and the slide was neat; sufficient epithelial cells were present; no, or only scattered, poorly interpretable clumps of cells were found; and there were no signs of inflammation, sufficient to hinder assessment. A smear was considered pathological when at least mild dysplasia was found (pap ⩾IIIa).

F Buntinx, MD, PhD, general practitioner and researcher; J A Knottnerus, MD, PhD, professor; H F J M Crebolder, MD, PhD, professor; T Seegers, research assistant, Department of General Practice, State University of Limburg, The Netherlands. G G M Essed, MD, PhD, professor, Department of Gynaecology, State University of Limburg, The Netherlands. H Schouten, PhD, associate professor, Department of Medical Statistics, State University of Limburg, The Netherlands.
Submitted: 16 March 1992; accepted: 3 August 1992.

Determination of baseline

During the first three months of assessment the quality of each smear was registered but the doctors continued to receive the usual laboratory report with a description of the cells and the pap-classification for each smear submitted. The data from this period were used to determine a baseline, with which the results of the intervention period could be compared.

Qualification period

The next two months were considered to be a qualification period.[10] During this period all 183 study doctors were provided with a copy of an article with photographs on the correct technique for obtaining cervical smears with different instruments.

Intervention

A six month intervention period followed the qualification period. At the beginning of the intervention period the 183 doctors were randomly assigned to one of four intervention groups.

Group A. No change took place, and this group served as a control group.

Group B. A comment based on the results of the quality assessment was added to every laboratory report: 'The technical quality of this smear was excellent/moderate/unsatisfactory.' The reason why smears were assessed as 'unsatisfactory' was also given.

Group C. Identical to group B, but in addition, doctors were provided with an overview each month of the quality assessments of the smears submitted by them in the previous month (percentage of smears rated as excellent, moderate and unsatisfactory), compared with the mean for the entire group (peer comparison).

Group D. Identical to group C, but these doctors also received, when required, specific advice concerning deficiencies in their technique together with information about relevant methods of postgraduate education.

The randomization was stratified by discipline (general practitioner or gynaecologist), by the number of smears submitted in 1988 and by the usual sampling technique, and was based on computer generated random numbers.

Sample size

In order to be able to determine a reduction of at least 5% in the mean percentage of unsatisfactory smears for each doctor relative to the control group, 20 doctors would be required in each group, assuming an alpha error of 0.05, a beta error of 0.05 and a standard deviation of 4%. This number was achieved for all of the intervention groups, and for some of the subgroups.

Analysis

The quality assessments for each smear and for each doctor were collected for each intervention group and period.

Crude data analysis. For each intervention group, the mean of the percentage of inadequate smears for each doctor in the baseline and intervention periods was calculated. These means were the basis for further analyses, together with the mean percentages of pathological smears.

For each doctor the difference between the percentage of inadequately sampled smears during the intervention period and during the baseline period (intervention difference), was calculated, together with a quotient for these two percentages (intervention quotient). An intervention quotient of less than one or an intervention difference of less than zero indicate a decrease in the number of smears of unsatisfactory quality for a particular criterion.

For each group the means of these parameters were calculated together with a 95% confidence interval (non-parametrically calculated). The mean intervention differences and the intervention quotients of the intervention groups and control group were compared. Since it could not be assumed that the results would have a normal distribution, non-parametric tests for significance were used (Mann Whitney and Kruskall Wallis). In order to determine whether there was a linear relationship between the intervention difference or intervention quotient and intensity of the intervention a Spearman rank correlation coefficient was calculated.[11] A two-sided significance level of 5% was chosen in statistical testing. Computations were carried out with the *BMDP* programme.[12]

During the baseline period, a number of doctors in every group achieved an adequate result for a particular quality indicator for all smears submitted. For these doctors, an improvement was not possible, and the provision of feedback less relevant. In addition, it was not possible to compute an intervention quotient for these doctors since the denominator should be zero.

Subgroup analysis. The analysis was repeated for each discipline and sex of the doctors, for each category of sampling method used (this was relevant for the absence of endocervical cells only) and for three categories of the number of smears submitted during the intervention period.

Regression analysis. For the presence of endocervical cells — the most relevant characteristic of quality — the effect of the interventions was analysed, adjusted for a number of covariables by forward stepwise multiple regression. For the other quality indicators such an analysis was not feasible owing to the paucity of data remaining after the required logarithmic transformation (the logarithm of zero cannot be calculated).

Analysing the influence of each newly introduced intervention technique, the effect of the additional statement on the laboratory report (groups B, C and D versus group A), of the monthly overviews with peer review (groups C and D versus groups A and B) and of the specific advice (group D versus groups A, B and C) were computed.

The dependent variable was the natural logarithm of the intervention quotient. The independent variables were the intervention group, the percentage of smears lacking endocervical cells during the baseline period, the sampling technique employed (using four dummy variables), the sex and discipline of the doctor, and the total number of smears submitted during the intervention period. The dummy variables for the sampling method were introduced or removed as a set from the regression equation. Starting from an empty model the required P value for introduction or removal of a variable was 0.05.

Results

Four of the 183 doctors were excluded from the study after randomization; two from the control group, and one each from groups C and D. One of the four doctors was found to be aware of the study, another received feedback although he was assigned to the control group, and two did not submit any smears during the intervention period. For a further two doctors the professional data were incomplete. Thus, the majority of the analyses reported here concern 177 doctors.

The 179 doctors participating in the trial were distributed throughout the five Flemish provinces of Belgium. The mean age of the doctors was 39 years (standard deviation 13 years) and 77.7% were men. No statistically significant differences between

the four intervention groups with respect to these variables were observed. The distribution among the four cytotechnologists of the 17 664 smears submitted by the 179 doctors over the 11 month period was as follows: 29.2%, 28.4%, 19.9% and 22.5%, again with no significant differences between the four intervention groups. In addition, the intervention groups were comparable with respect to discipline and sampling technique employed (Table 1).

Table 1. Sampling technique employed and number of smears taken by the participating doctors.

	Intervention group			
	A	B	C	D
Number of doctors				
Total	46	46	43	44
GPs	42	40	41	41
Gynaecologists	4	6	2	3
Sampling technique (number of doctors)[a]				
Cytobrush®, with or without spatula	26	25	23	22
Spatula (single or combined)	13	15	18	17
Miscellaneous	6	5	2	4
Number of smears[b]				
Total	4805	5863	3725	3273
Mean for all doctors	104.5	127.5	86.6	74.4
Mean for GPs	59	71	65	57

[a]Data missing for one doctor. [b]Over the 11 month study period.

Crude data analysis

During the baseline period less than 1% of the smears in each intervention group scored unsatisfactorily for the criteria: presence of sufficient epithelial cells, poorly interpretable clumps of cells and signs of inflammation. Therefore, no clear improvement would be possible, and the influence of the interventions on these indicators was not investigated any further.

The interventions were found to have no systematic influence on the presence of endocervical cells or on the quality of fixation (Table 2). An increase in the percentage of smears containing pathological cells with the intensity of the intervention was noted. However, zero lies within the 95% confidence interval for the intervention difference for every group and the Mann Whitney test comparing the different intervention groups with the control group also gave non-significant results. In addition, the correlation coefficients are all close to zero and non-significant, thus failing to demonstrate a clear correlation between the intensity of the intervention and the increase in improvement in all three respects.

Subgroup analysis and regression analysis

A systematic decrease in the number of smears lacking endocervical cells was associated with more intensive feedback among gynaecologists and those doctors who submitted more than 50 smears during the intervention period (Table 3). The correlation between intervention quotient and increasing intervention was significant for these two groups. The Kruskall Wallis test for the intervention quotients was significant for the doctors who submitted more than 50 smears ($P<0.05$). A similar pattern was found for intervention differences. Using multiple regression analysis with adjustment for the covariables mentioned above, no significant result was found when groups B, C and D combined

Table 2. Percentage of poor quality smears and those containing pathological cells, by intervention group.

Smears with:	Intervention group A (n = 45)	B (n = 46)	C (n = 43)	D (n = 43)	Spearman rank correlation coefficient
Absence of endocervical cells					
Mean % of smears during					
Baseline period	*20*	*17*	*15*	*19*	
Intervention period	*15*	*12*	*8*	*11*	
Mean intervention difference (%) (95% CI)	*−9 (−13 to −5)*	*−10 (−16 to −5)*	*−13 (−19 to −8)*	*−12 (−16 to −7)*	−0.11
Mean intervention quotient (95% CI)	0.69 (0.53 to 0.85)	0.69 (0.47 to 0.91)	0.46** (0.26 to 0.67)	0.62 (0.44 to 0.79)	−0.15
[no. of doctors[a]]	[33]	[34]	[29]	[34]	
Inadequate fixation					
Mean % of smears during					
Baseline period	*7*	*3*	*2*	*3*	
Intervention period	*0.5*	*2*	*1*	*1*	
Mean intervention difference (%) (95% CI)	*−24 (−45 to −3)*	*−4 (−10 to −2)*	*−9 (−15 to −2)*	*−10 (−13 to −6)*	−0.07
Mean intervention quotient (95% CI)	0.16 (−0.46 to 0.33)	0.63* (0.12 to 1.15)	0.09 (0.00 to 0.20)	0.09 (0.00 to 0.20)	−0.12
[no. of doctors[a]]	[13]	[15]	[8]	[11]	
Pathological cells[b]					
Mean % of smears during					
Baseline period	*0.54*	*0.35*	*0.49*	*0.61*	
Intervention period	*0.59*	*0.61*	*0.81*	*1.06*	
Mean intervention difference (%) (95% CI)	*0.06 (−0.79 to 0.90)*	*0.25 (−0.31 to 0.81)*	*0.31 (−0.32 to 0.95)*	*0.44 (−0.53 to 1.40)*	0.01
Mean intervention quotient	_[c]	_[c]	_[c]	_[c]	

n = number of doctors in group. CI = confidence interval. [a]Eligible for the calculation of the intervention quotient. [b]n = 45, 46 42, 42, respectively. [c]Too few doctors to allow calculation. *$P<0.05$; **$P<0.01$ (Mann Whitney test for the difference between intervention group and control group A).

were compared with group A or when group D was compared with groups A, B and C combined. However, the odds ratio for groups C and D combined versus groups A and B combined (the effect of the monthly overviews and peer comparison) is 0.75 (95% confidence interval 0.58 to 0.96). The Cook distance here is maximally 0.05, which indicates the absence of extremes.

No systematic improvement in the quality of fixation was found in any subgroup. However, most doctors submitted only well-fixed smears during the baseline period (Table 2).

For the analysis of percentage of smears containing pathological cells only the intervention differences were examined, owing to the small number of doctors who could be included in the analysis for intervention quotients. It appears that the non-significant but systematic increase in the percentage of abnormal smears is attributable to the general practitioners; it is not found in any of the other subgroups (Table 4).

Discussion

This was a double blind randomized trial: the cytotechnologists assessing quality were not aware of which intervention group the doctors had been allocated to and the doctors were not aware of being studied. Thus it was possible to exclude any observation bias or Hawthorne effect.

In the crude data analysis the intervention was found to have no systematic influence on the mean intervention quotients or differences for the quality characteristics examined. There was, however, a noticeable increase in the percentage of smears with pathological cells with increasing intensity of intervention. This increase was not statistically significant but this may be attributable to the low prevalence of pathology. To demonstrate a statistically significant increase in the intervention effect of 0.5% (or 1%) with an alpha error of 10% and a beta error of 10%, would require 2373 (or 739) doctors in each intervention group.

In the subgroup analysis, it appeared that the increase in the number of abnormal smears with intensity of intervention could be virtually entirely assigned to the subgroup of general practitioners. It would be interesting to test this observation with a larger group but this would appear possible only by means of a multicentre study in which a large number of laboratories introduced comparable feedback methods simultaneously.

Although the crude data analysis did not reveal any effect of feedback on the presence of endocervical cells in cervical smears, a stepwise multiple regression analysis yielded a relative decrease in the groups which received monthly overviews and peer comparison, which was one third bigger than in the other groups (odds ratio 0.75). This suggests a limited but demonstrable effect of this type of feedback on this quality indicator.

The analysis of the pattern of improvement for the absence of endocervical cells in the various subgroups yields a rather complex picture. On closer examination, however, it can be seen that an improvement as a result of feedback is noticeable in those subgroups in which the doctors produce most smears over a period of time — gynaecologists, and doctors submitting more than 50 smears during the intervention period. These doctors receive

Table 3. Influence of the intervention on the percentage of smears with no endocervical cells.

Absence of endocervical cells	Mean intervention quotient (no. of doctors[a]) for intervention group				Spearman rank correlation coefficient
	A	B	C	D	
GPs	0.67 (30)	0.67 (28)	0.48 (27)	0.63 (31)	-0.11
Gynaecologists	0.87 (3)	0.79 (6)	0.29 (2)	0.42 (3)	-0.59*
Mean number of smears submitted during intervention period					
⩽ 25	0.63 (13)	0.68 (13)	0.37* (7)	0.73 (16)	-0.05
26–50	0.68 (9)	0.59 (7)	0.51 (12)	0.66 (10)	0.02
>50	0.77 (11)	0.75 (14)	0.47 (10)	0.35** (8)	-0.44**

[a] Eligible for the calculation of the intervention quotient. *$P<0.05$; **$P<0.01$ (Mann Whitney test for the difference between intervention group and control group A; and the correlation coefficient).

Table 4. Influence of the intervention on the percentage of smears showing at least mild dysplasia.

Smears containing pathological cells	Mean intervention difference (%) [no. of doctors] (95% CI) for intervention group				Spearman rank correlation coefficient
	A	B	C	D	
GPs	*-0.05* [42] (*-0.93* to *0.83*)	*0.34* [40] (*-0.28* to *0.97*)	*0.25* [40] (*-0.41* to *0.92*)	*0.43* [39] (*-0.61* to *1.48*)	-0.01
Gynaecologists	*1.62* [3] (*-4.21* to *7.44*)	*-0.35* [6] (*-1.68* to *0.99*)	*1.57* [2] (*-3.87* to *7.01*)	*0.49* [3] (*-1.62* to *2.60*)	-0.05
Number of smears submitted during intervention period					
⩽ 25	*0* [22]	*-0.24* [19] (*-0.74* to *0.26*)	*-0.03* [15] (*-1.30* to *1.25*)	*0.99* [21] (*-0.41* to *2.38*)	0.13
26–50	*-0.98* [12] (*-4.00* to *2.00*)	*0.26* [11] (*-0.83* to *1.36*)	*0.45* [14] (*-0.52* to *1.41*)	*0.13* [13] (*-1.43* to *1.69*)	0.05
>50	*1.31* [11] (*-0.23* to *2.85*)	*0.83* [16] (*-0.57* to *2.23*)	*0.56* [13] (*-0.78* to *1.90*)	*-0.51* [8] (*-3.87* to *2.86*)	-0.14

CI = confidence interval.

feedback most often as they submit smears virtually daily. Thus whatever they learn from the feedback may be applied immediately. Doctors who produce few smears receive feedback rarely and may ascribe poor results in the monthly overviews to chance, thus learning little from the feedback. In addition, they may not perform another smear for several days when good intentions may have been forgotten. It is possible that those doctors who submit few smears are less interested in cervical screening and have less motivation to achieve good results.

The results of this study suggest that feedback, in the form of written quality assessments plus monthly overviews and peer comparison, has a positive influence on the presence of endocervical cells in the smears submitted. This is especially true when the number of smears is sufficient to allow the feedback to have an effect — the lower limit might be 100 smears per year. This study has been carried out among a group of practitioners whose baseline performance was 15–20% of smears lacking endocervical cells. These figures are lower than has been found by other authors in Belgium, the United Kingdom and elsewhere.[13] It is possible that the effect of this type of feedback on doctors with poorer initial results would be even more obvious.

The intervention had no effect on the quality of fixation. However, many doctors submitted only well-fixed smears during the baseline period, leaving no room for improvement and thereby rendering these subjects ineligible for analysis. Consequently, the power of this part of the analysis is greatly reduced.

This intervention is inexpensive and not labour intensive. In any system it is important that the effect of the intervention is not primarily based on the personal efforts of a programme manager.[14] The system described here could be adopted by any laboratory or screening agency and was welcomed by the doctors that participated.[15] Feedback is increasingly used as a means of improving the performance of medical practitioners. However, since there have been few carefully performed studies comparing various methods,[16–18] there remains a need for controlled trials on the effects of different types of feedback.

References

1. Mitchell H, Medley G. Longitudinal study of women with negative cervical smears according to endocervical status. *Lancet* 1991; **338:** 265-267.
2. Kivlahan C, Ingram E. Papanicolaou smears without endocervical cells. Are they inadequate? *Acta Cytol* 1986; **30:** 258-260.
3. Vooijs GP. Significance of cellular composition of smears for the reliability of cytological diagnosis. In: Goerttler K, Fiechter G, Witte S (eds). *New frontiers in cytology.* Berlin, Germany: Springer-Verlag, 1988.
4. Mauney M, Eide D, Sotham J. Rates of condyloma and dysplasia in papanicolaou smears with and without endocervical cells. *Diagn Cytopathol* 1990; **6:** 18-21.
5. Buntinx F, Boon ME, Beck S, *et al.* Comparison of Cytobrush sampling, spatula sampling and combined Cytobrush-spatula sampling of the uterine cervix. *Acta Cytol* 1991; **35:** 64-68.
6. Buntinx F, Knottnerus A, Crebolder HFJM, *et al.* Relation between quality of cervical smears and probability of abnormal results. *BMJ* 1992; **304:** 1224.
7. Buntinx F, Crebolder HFJM, Knottnerus JA, Essed GGM. Cervical smear sampling quality. *Lancet* 1991; **337:** 926-927.
8. Boon ME, Alons-van Kordelaar JJM, Rietveld-Scheffers PEM. Consequences of the combined spatula and Cytobrush sampling for cervical pathology. *Acta Cytol* 1986; **30:** 264-270.
9. Baker RM. Improving the adequacy of pap smears. *Am Fam Physician* 1989; **39:** 109-114.
10. Knipschild P, Leffers P, Feinstein AR. The qualification period. *J Clin Epidemiol* 1991; **44:** 461-464.
11. Armitage P, Berney G. *Statistical methods in medical research.* 2nd edition. Oxford: Blackwell, 1987.
12. Dixon WJ, Brown MB, Enselman L, Jennrich RI. *BMDP — statistical software.* Los Angeles, CA: University of California Press, 1990.
13. Buntinx F, Essed GGM, Crebolder HFJM, Knottnerus JA. Geographical variations in the sampling quality of cervical smears. *Eur J Obstet Gynecol Reprod Biol* 1991; **42:** 227-230.
14. Marton KI, Tul V, Sox HC. Modifying test-ordering behavior in the outpatient medical clinic — a controlled trial of two educational interventions. *Arch Intern Med* 1985; **145:** 816-821.
15. Buntinx F, Knottnerus JA, Crebolder HFJM, Essed GGM. Reactions of doctors to various forms of feedback designed to improve the sampling quality of cervical smears. *Quality Assurance in Health Care* 1992; **4:** 161-166.
16. Eisenberg JM, Williams SV. Cost containment and changing physicians' practice behavior. Can the fox learn to guard the chicken coop? *JAMA* 1982; **246:** 2195-2201.
17. Fraser CG, Woodford FP. Strategies to modify the test-requesting patterns of clinicians. *Ann Clin Biochem* 1987; **24:** 223-231.
18. Mitchell MW, Fowkes FGR. Audit reviewed: does feedback on performance change clinical behaviour? *J R Coll Physicians Lond* 1985; **19:** 251-254.

Acknowledgements

This study was was made possible by the financial support of the Praeventiefonds (project 28-1872) and the Flemish Institute of General Practice. We thank the staff of the department of cytology, Central Laboratory, Antwerp.

Address for correspondence

Dr F Buntinx, Dr Haubenlaan 43, B 3630 Maasmechelen, Belgium.