

from InterStudy (now called the Health Outcomes Institute) in America, on which the anglicised versions used in the Sheffield study and the Oxford healthy life study were based, and two more recent versions. One of these available from John Ware at the New England Medical Center Hospital<sup>2</sup> and the other from the Rand Corporation.<sup>23</sup> The latter versions have thus far not been validated for use in England and, despite the relatively minor changes between these and the original questionnaire, we would suggest that users continue to use the original InterStudy version, adapted for use in England.

We thank Sue Bradshaw, Peter Brooks, and Diana Harwood for help with the administration of the survey, Sue Ziebland for comments on an earlier draft of the paper, Berkshire, Buckinghamshire, Northamptonshire, and Oxfordshire Family Health Services Authorities for help in drawing the sample, and Oxford Regional Health Authority for funding the study.

- 1 Geigle R, Jones S. Outcomes measurement: a report from the front. *Inquiry* 1990;27:7-13.
- 2 Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36) I: conceptual framework and item selection. *Med Care* 1992;30:473-83.
- 3 Brazier JE, Harper R, Jones NMB, O'Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992;305:160-4.
- 4 Wilkin D, Hallam L, Doggett M. *Measures of need and outcome for primary care*. Oxford: Oxford University Press, 1992.
- 5 McDowell I, Newell C. *Measuring health: a guide to rating scales and questionnaires*. New York: Oxford University Press, 1987.
- 6 Bowling A. *Measuring health: a review of quality of life measurement scales*. Milton Keynes: Open University Press, 1991.

- 7 Hunt S, McEwen J, McKenna SP. Measuring health status: a new tool for clinicians and epidemiologists. *JR Coll Gen Pract* 1985;35:185-8.
- 8 Hunt S, McEwen J, McKenna SP. *Measuring health status*. London: Croom Helm, 1986.
- 9 Kind P, Carr-Hill R. The Nottingham health profile: a useful tool for epidemiologists? *Soc Sci Med* 1987;25:905-10.
- 10 Ware JE, Brook RH, Williams KN, Stewart AL, Davies-Avery A. *Conceptualisation and measurement of health for adults in the health insurance study*. Vol 1. Model of health and methodology. Santa Monica, California: Rand Corp, 1980.
- 11 Stewart AL, Ware JE, eds. *Measuring functioning and well being: the medical outcomes study approach*. London: Duke University Press, 1992.
- 12 Stewart AL, Hays RD, Ware JE. Methods of validating the MOS health measures. In: Stewart AL, Ware JE, eds. *Measuring functioning and well being: the medical outcomes study approach*. London: Duke University Press, 1992.
- 13 Wright L, Harwood D, Coulter A. *Health and lifestyles in the Oxford region*. Oxford: Health Services Research Unit, 1992.
- 14 Office of Population Censuses and Surveys. *Standard occupational classification*. Vol 2. London: HMSO, 1990.
- 15 Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297-334.
- 16 Helmstadter GC. *Principles of psychological measurement*. New York: Appleton-Century-Crofts, 1964.
- 17 Nunnally JC. *Psychometric theory*. 2nd ed. New York: McGraw Hill, 1978.
- 18 Carmines E, Zeller R. *Reliability and validity assessment. Quantitative applications in the social science*. Beverly Hills: Sage, 1979.
- 19 Bergner M, Bobbitt RA, Carter WB, Gilson BS. The sickness impact profile: development and final revision of a health status measure. *Med Care* 1981;19:787-805.
- 20 Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. The importance of sensitivity to change as a criterion for selecting health status measures. *Quality in Health Care* 1992;1:89-93.
- 21 Ziebland S, Fitzpatrick R, Jenkinson C. Assessing short term outcome. *Quality in Health Care* 1992;1(2):141-2. (Letter.)
- 22 Carr-Hill R, Morris J. Current practice in obtaining the "Q" in QALYs: a cautionary note. *BMJ* 1991;303:699-701.
- 23 Rand Health Sciences Program. *Rand 36-item health survey 1.0*. Santa Monica: Rand, 1992.

(Accepted 16 March 1993)

## The SF 36 health survey questionnaire: an outcome measure suitable for routine use within the NHS?

Andrew M Garratt, Danny A Ruta, Mona I Abdalla, J Kenneth Buckingham, Ian T Russell

### Abstract

**Objective**—To assess the validity, reliability, and acceptability of the short form 36 (SF36) health survey questionnaire (a shortened version of a battery of 149 health status questions) as a measure of patient outcome in a broad sample of patients suffering from four common clinical conditions.

**Design**—Postal questionnaire, followed up by two reminders at two week intervals.

**Setting**—Clinics and four training practices in north east Scotland.

**Subjects**—Over 1700 patients aged 16-86 with one of four conditions—low back pain, menorrhagia, suspected peptic ulcer, or varicose veins—and a comparison sample of 900 members of the general population.

**Main outcome measures**—The eight scales within the SF36 health profile.

**Results**—The response rate exceeded 75% in the patient population (1310 respondents). The SF36 satisfied rigorous psychometric criteria for validity and internal consistency. Clinical validity was shown by the distinctive profiles generated for each condition, each of which differed from that in the general population in a predictable manner. Furthermore, SF36 scores were lower in referred patients than in patients not referred and were closely related to general practitioners' perceptions of severity.

**Conclusions**—These results provide support for the SF36 as a potential measure of patient outcome within the NHS. The SF36 seems acceptable to patients, internally consistent, and a valid measure of the health status of a wide range of patients. Before it can be used in the new health service,

however, its sensitivity to changes in health status over time must also be tested.

### Introduction

The government has responded to the need for efficient provision of health care by introducing managed competition into the National Health Service.<sup>1</sup> In the internal market health authorities and fundholding practices may purchase care from competing provider units. If the reforms are to be successful, then purchasers will require valid, reliable, and sensitive measures of outcome to allow them to allocate scarce resources in the most cost effective manner. Similarly, service providers who can demonstrate the effectiveness of the care they provide in improving patient outcome will be better placed to compete for purchaser funds. Without such information health care will be purchased on the basis of cost alone, with serious consequences for its quality.

Few outcome measures currently available for routine use satisfy the criteria of validity, reliability, and sensitivity to changes in health status. For example, the Nottingham health profile has been criticised for failing to detect low levels of morbidity.<sup>2</sup> A measure that deserves careful consideration is the short form 36 (SF36) health survey questionnaire.<sup>3,4</sup> The SF36 is a shortened version of a battery of 149 health status questions developed and tested on a population of over 22 000 patients as part of the medical outcome study,<sup>5,6</sup> designed to help understand how specific components of the American health care system affect the outcomes of care. A key objective of the study was to develop more practical tools for monitoring patient outcomes in

Department of Public Health, University of Aberdeen, Foresterhill, Aberdeen AB9 2ZD  
Andrew M Garratt, research assistant  
Danny A Ruta, lecturer

Health Services Research Unit  
Mona I Abdalla, research fellow  
J Kenneth Buckingham, research fellow  
Ian T Russell, director

Correspondence to: Mr Garratt.

BMJ 1993;306:1440-4

a busy clinical setting. Questions were selected to produce a questionnaire that could be completed in under 10 minutes while retaining the validity and reliability of the longer parent questionnaire.

The SF 36 is a general outcome measure. In contrast with condition specific measures, which may be criticised for their narrowness, general measures attempt to capture aspects of health that are important to all patients. They are useful for health status comparisons both among patients with the same condition and between patients with different conditions. Such measures can also be administered to general populations to see how a particular condition causes health to depart from a "healthy standard."<sup>7,8</sup> The SF 36 uses eight health scales to measure three aspects of health—functional status, wellbeing, and "overall evaluation of health" (see table I). The responses to the questions on each scale are summed to provide eight scores between 0 and 100.

TABLE I—Internal consistency of SF 36 health profile

SF 36 scales	No of items	Internal consistency (Cronbach's $\alpha$ )
I. Functional status:		
(a) Physical functioning	10	0.92
(b) Social functioning	2	0.80
(c) Role limitations attributed to physical problems	4	0.89
(d) Role limitations attributed to emotional problems	3	0.86
II. Wellbeing:		
(a) Mental health	5	0.86
(b) Energy and fatigue	4	0.86
(c) Pain	2	0.86
III. Overall evaluation of health:		
(a) General health perception	5	0.83
Total	35*	

\*36th question (asking respondents to compare present health with that one year before) not included within the eight scales.

The validity and reliability of the SF 36 in patient populations has been confirmed in the United States.<sup>4,9</sup> Patients were classified by the severity of their medical conditions. For example, patients with chronic heart failure reporting oedema, orthopnoea, or dyspnoea on exertion were classified as having a serious medical condition. The SF 36 could detect differences in health status among these patient groups across all eight scales. In particular, it was possible to distinguish patients with medical conditions from those with psychiatric conditions. Moreover, by examining the distinctive profiles of the SF 36 scores across the eight scales it was possible to distinguish patients with mixed chronic medical and psychiatric conditions in which the psychiatric component was predominant from those in which the medical component was predominant.

These and similar findings provide strong evidence for the clinical validity of the SF 36 as a measure of patients' perceived health. If the SF 36 is to be adopted as a measure of outcome by the NHS these results will have to be replicated in a wide range of British populations. Preliminary findings from a general population are encouraging,<sup>10</sup> but no study has looked at a British patient population. We report the results of a study in which the SF 36 was administered to a large sample of patients suffering from four common conditions—low back pain, menorrhagia, suspected peptic ulcer, and varicose veins.

### Subjects and methods

Between March and June 1991 we identified patients in Grampian presenting with one of four common conditions—low back pain, menorrhagia, suspected peptic ulcer, or varicose veins. These patients were identified in one of two ways—from all referral letters

to outpatient departments in Grampian, and by general practitioners from four large training practices in Grampian; this second group of patients was included only if the general practitioner did not refer them to a specialist during the recruitment period. General practitioners were also asked to assess their patients on a four point scale of symptom severity (none, mild, moderate, severe).

A questionnaire including an anglicised version of the SF 36<sup>10a</sup> and sociodemographic questions was sent to the patients in general practice within two weeks of their initial consultation and to referred patients before their first outpatient appointment. Patients not wishing to take part in the study were asked to return their questionnaires blank. Reminders were sent to non-respondents after two weeks and again after four weeks. A random sample of 900 members of the general population, selected from the electoral register for Aberdeen, served as a comparison group. They were sent a similar questionnaire.

To test for response bias, patients who did not respond or refused to take part in the study were compared with respondents in age, gender, clinical condition, source, and reported symptom severity. In addition, a sample of non-respondents were telephoned at home and asked questions about health and socio-economic status. Detailed analysis of these interviews will be reported later.

### STATISTICAL ANALYSIS

An outcome measure should be both reliable and valid. Validity is the extent to which an instrument measures what is intended.<sup>11</sup> Reliability is the extent to which similar measurements on the same person are similar in different settings.<sup>11</sup> In this paper we assess reliability through internal consistency, which measures the extent to which similar questions produce consistent responses.<sup>11</sup>

### INTERNAL CONSISTENCY

The SF 36 would be internally consistent if the responses to items (that is, questions) that contribute to the same scale correlate well with each other. Two techniques were used to assess the internal consistency of the SF 36—item-scale correlations<sup>11</sup> and Cronbach's  $\alpha$ .<sup>12</sup> Item-scale correlations, which assess the extent to which an item is related to the remainder of its scale, should exceed 0.4.<sup>13</sup> Furthermore, items should be more closely related to their own scale than the other scales. Cronbach's  $\alpha$  measures the overall correlation between items within a scale. Reliability is considered acceptable for group comparisons when  $\alpha$  exceeds 0.7.<sup>14</sup>

### VALIDITY

Tests of validity can be divided into tests based on psychometric criteria and tests based on clinical and social criteria. Confirmatory factor analysis,<sup>15</sup> a technique of psychometric validation, assesses the agreement between hypothetical factors that go to make up the measure and the scales designed to assess those factors. If the SF 36 is a valid measure for use in Britain the scales defined by its authors (table I) should emerge from a factor analysis of British patient data, and items relating to a particular scale should be grouped together within a single factor. Within such an assessment a factor should be considered relevant only if its "eigenvalue" (a statistical measure of its power to explain variation between patients) exceeds 1.1.<sup>16</sup>

Construct validity assesses the extent to which a measure is related to criteria derived from an established clinical or social theory or "construct." In assessing the SF 36 for clinical construct validity we hypothesised that scores should vary in a predictable manner among the four clinical conditions and the

general population. For example, the largest differences from the general population were expected in patients with back pain and the smallest in patients with varicose veins. We also predicted that referred patients would report worse health than non-referred patients, and that SF 36 scores would correspond to general practitioners' perceptions of symptom severity.

Ordinary least squares regression was used to estimate the effect on each of the eight SF 36 health scales of each of the four conditions; age, gender, and socioeconomic status; whether the patient was referred; and symptom severity as reported by the general practitioner. The distributions of the scores for some of the scales within the SF 36 and the closely related SF 20 are often skewed.<sup>17,18</sup> We therefore checked both the distributions of the eight scales and, even more important, the residual plots from the corresponding regression analyses.<sup>19</sup>

The results are presented as absolute and standard scores, both for a member of the general population with average characteristics and for members of the four patient groups with the same characteristics. Absolute scores, presented as mean deviations from the scores of the general population, allow comparisons between the patient groups and the general population for each individual SF 36 scale. Standard scores are calculated by dividing the differences between the scores of each condition specific patient group and those of the general population by the standard deviation of the general population. Presented as line graphs, these allow comparisons between the patient groups and the general population across the entire SF 36 health profile.

TABLE II—Factor analysis

Factor coefficients of individual questions after rotation					
	F1	F2	F3	F4	F5
	<i>Physical</i>	<i>Mental</i>	<i>Role—physical</i>	<i>General</i>	<i>Role—emotional</i>
3a	0.55	9b 0.40	4a 0.76	1 0.74	5a 0.75
3b	0.68	9d 0.70	4b 0.73	10a 0.71	5b 0.78
3c	0.69	9f 0.61	4c 0.78	10b 0.77	5c 0.75
3d	0.74	9h 0.70	4d 0.75	10c 0.63	
3e	0.80	9c 0.60		10d 0.78	
3f	0.64		<i>Social</i>		
3g	0.79	<i>Energy</i>	6 0.52		
3h	0.82	9a 0.73	9j 0.44		
3i	0.72	9e 0.76			
3j	0.58	9g 0.70	<i>Pain</i>		
		9i 0.74	7 0.63		
			8 0.71		
Eigenvalue before rotation*	12.8	3.8	2.1	1.8	1.3

\*Statistical measure of power to explain variation between patients.<sup>16</sup>

## Results

### RESPONSE RATE

In total 1787 patients were identified, 800 by their general practitioners and 987 from referral letters. Of these, 236 failed to respond, 193 refused to take part, and 41 questionnaires were returned by the Post Office. Thus 1317 of 1746 correctly identified patients took part, giving a final response rate of 75.5%; 780 patients (44.6% of 1746) responded before the first and 1150 (65.9%) before the second reminder. All but nine patients attempted the SF 36; their ages ranged from 16 to 86 (mean 42.7 years), and 870 (66.5%) were women. The 429 patients who did not respond were significantly younger (mean age 39.9 years;  $p < 0.01$ ). However, they did not differ significantly from respondents in gender, clinical condition, source, or symptom severity as reported by their general practitioners. Of the comparison sample of 900 members of the general population, 542 (60.2%) returned a questionnaire. Their ages ranged from 18 to 91 (mean 47.9 years), and 292 (53.9%) were women.

### INTERNAL CONSISTENCY

The item total correlation coefficients for the 35 items within the eight health scales ranged from 0.55 to 0.78, thus satisfying Kline's criterion of 0.4.<sup>13</sup> The correlation coefficients between items and the remainder of their own scale were all higher than those with other scales, providing further evidence of internal consistency. For all eight scales internal consistency measured by Cronbach's  $\alpha$  exceeded 0.80, thus satisfying Nunnally's criterion of 0.7<sup>14</sup> (table I).

### VALIDITY

Factor analysis identified five relevant factors, with eigenvalues ranging from 12.8 to 1.3 (table II). No health scale was spread across more than a single factor. The first factor (F1) represents physical functioning; the second (F2) represents mental health and energy; the third (F3) represents social functioning, pain, and role limitations attributable to physical problems; the fourth (F4) represents general health perception; and the fifth (F5) represents role limitations attributable to emotional problems. This precise correspondence between factors and scales is rare in factor analysis<sup>20</sup> and thus particularly reassuring for the validity of the SF 36 in British patients.

Table III shows the mean scores for the general population and the mean deviations for condition specific groups corrected for age, gender, and socioeconomic status. Back pain patients scored less than

TABLE III—Mean SF 36 scores for general population and mean deviations for condition specific groups (corrected for differences in age, sex, and socioeconomic status)

Group	No of subjects	Physical functioning	Social functioning	Role—physical	Role—emotional	Mental health	Energy and fatigue	Pain	General health perception
General population	542	79.2	78.6	76.5	75.0	73.7	61.2	76.9	68.7
Patients suffering from:									
Low back pain	558	-26.7**	-20.8**	-56.0**	-29.6**	-12.2**	-18.7**	-42.6**	-12.1**
Menorrhagia	271	-1.6	-12.1**	-29.7**	-26.0**	-14.3**	-21.3**	-23.5**	-8.7**
Suspected peptic ulcer	203	1.7	-7.0**	-16.2**	-9.6**	-9.1**	-12.3**	-22.9**	-11.5**
Varicose veins	278	-5.3**	0.2	-11.1**	-8.4**	-3.1*	-4.9**	-8.5**	-1.3
Referred patients suffering from:									
Low back pain	244	-36.6**	-26.6**	-62.9**	-39.9**	-17.3**	-24.5**	-47.3**	-17.6**
Menorrhagia	200	-0.6	-13.0**	-31.1**	-27.5**	-15.1**	-22.9**	-24.8**	-9.4**
Suspected peptic ulcer	98	0	-10.4**	-22.9**	-13.0**	-12.9**	-18.6**	-28.1**	-15.5**
Varicose veins	220	-5.3**	0	-11.6**	-8.9**	-3.7*	-5.8**	-9.1**	-1.4
Non-referred patients suffering from:									
Low back pain	314	-18.6**	-16.0**	-50.4**	-21.3**	-8.1**	-13.9**	-38.8**	-7.5**
Menorrhagia	71	-2.6	-8.4**	-24.5**	-20.3**	-11.1**	-15.9**	-19.3**	-5.8**
Suspected peptic ulcer	105	3.5	-3.6	-9.6*	-6.0	-5.4**	-6.1**	-17.8**	-7.6**
Varicose veins	58	-4.2	1.9	-8.5	-4.9	0.2	-0.4	-5.6	-0.1
General practitioner severity ratings for non-referred patients†:									
None	15	3.0	3.0	-2.1	-4.0	-2.4	-3.4	3.0	6.7
Mild	214	-5.3	2.8	-10.4	-8.4	-4.9	-7.6	-2.8	4.7
Moderate	266	-9.7	-5.7	-20.9	-14.9	-10.2	-14.0	-11.2	-1.4
Severe	35	-21.0	-18.8	-34.1	-25.4	-15.2	-20.4	-16.9	-2.2

Compared with general population: \* $p < 0.05$ ; \*\* $p < 0.01$ .

†All eight SF 36 scales are linearly related to general practitioner severity ratings (F test for linear trend<sup>17</sup>;  $p < 0.01$ ).

Mean score for given patient group may be calculated by adding mean deviation to mean score for general population.

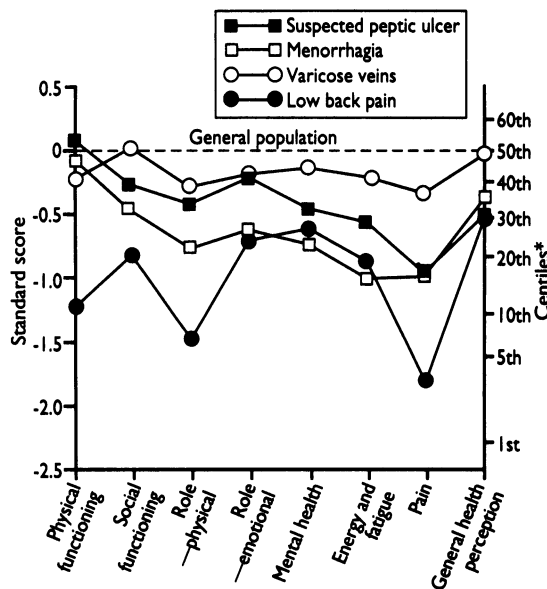


FIG 1—Mean SF36 health profiles for general population and patients with each of four study conditions  
\*Centiles of general population—for example, X% of general population falls below Xth centile and remaining members fall above that point.

other patients across five of the scales. The exceptions were mental health and energy or fatigue, where patients with menorrhagia scored as low as back pain patients, and general health perception, where peptic ulcer patients scored as low. Patients with varicose veins consistently scored the highest, the only exception being with respect to physical functioning. Only four of the differences in mean scores between patients and the general population failed to reach statistical significance—those for physical functioning in patients with menorrhagia and suspected peptic ulcer, and those for social functioning and general health perception in patients with varicose veins.

Figure 1 presents the standard scores for an average member of the general population and the same patient with each condition in turn. The mean for the general population is set at zero on each health scale, allowing comparisons to be made in terms of standard deviations of the general population. For example, low back pain patients with a mean standard score of about -2 for pain corresponded to the lowest scoring 2.5% of the general population. Each condition had a distinctive profile. Patients with varicose veins scored highest on seven of the eight health scales and had a health profile that was closest to that of the general population, followed by patients with suspected peptic ulcer, menorrhagia, and low back pain.

Referred patients consistently had lower average scores on all eight scales than non-referred patients with the same condition (table III, fig 2). The only exception was menorrhagia, where the physical functioning scores of both referred and non-referred patients were not significantly different from those of the general population.

Finally, table III shows that for all eight health scales mean scores were ordered strictly according to general practitioners' perceptions of symptom severity. This linear relation was significant at the 1% level for all eight scales.

## Discussion

To achieve its goals of efficiency and quality of care the new NHS needs an information system that will allow purchasers and providers to "estimate as best they can the relation between medical interventions and health outcomes."<sup>21</sup> Such a system will require measures of outcome that are valid, reliable, and

acceptable to patients. Our surveys have shown that the SF 36 health profile, developed and validated in the United States,<sup>3,7</sup> is acceptable to patients in general practice and outpatient clinics, with a response rate around 75%. Our findings also confirm that its psychometric validity and reliability have remained intact. For example, factor analysis confirms the relevance of the eight health scales proposed by the original authors.

Clinical validity was shown by clear differences in self reported health between the general population and patients with four common conditions. Highly significant differences were found even for varicose veins, often perceived as a minor condition. Further evidence of validity was provided by the differences that the SF 36 detected between referred and non-referred patients across all four conditions and all eight health scales, and by a high level of agreement between SF 36 scores and general practitioners' perceptions of symptom severity.

Together these last two findings suggest that, despite wide variations in referral patterns,<sup>22</sup> general practitioners base referral decisions, certainly in part, on the severity of patients' conditions. To what extent this severity is related to patients' capacity to benefit from specialist treatment we shall not know until we have analysed surveys that followed up our 1300 patients 12 months after their original referral or consultation. These surveys will also enable us to test the sensitivity of the SF 36 to changes in patients' health.

Finally, we emphasise that the SF 36 health profile is recommended not as a self contained questionnaire but as part of a more comprehensive portfolio of measures to assess many aspects of patient outcome. Such a portfolio should also include a condition specific measure of clinical outcome. As well as providing information that is more relevant to clinical decisions, such a measure might be more sensitive to changes in health after specialist treatment. The contribution of the SF 36 to such a portfolio of measures and its relation to clinically derived condition specific questionnaires<sup>20</sup> need further study. We hope that our follow up will contribute to the growing programme of British research into the SF 36 health profile.

We thank the staff at Inverurie, Portlethen, Rubislaw Place, and Westhill practices for recruiting patients; Jeremy Grimshaw, Jenny Duncan, and Alison De Ville for help with data collection; and John Ware and his colleagues at the Health Institute of the New England Medical Center for permission to use the SF 36 health profile. This research and the Health Services Research Unit are funded by the Chief Scientist Office of the Scottish Office Home and Health Department; however, the opinions expressed are ours alone.

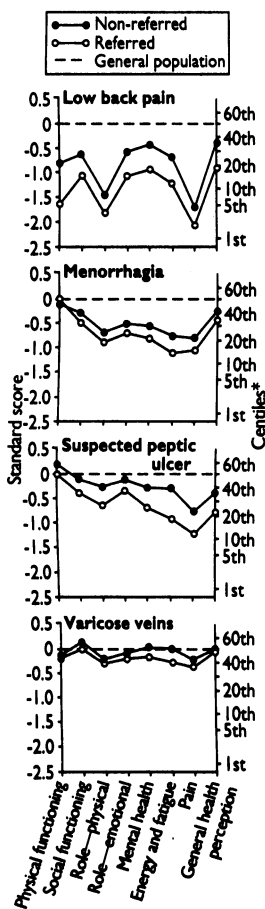


FIG 2—Mean SF36 health profiles for general population and non-referred and referred patients  
\*See footnote to fig 1.

- 1 Secretaries of State for Health, England and Wales, Northern Ireland, and Scotland. *Working for patients*. London: HMSO, 1989.
- 2 Kind P, Carr-Hill R. The Nottingham health profile: a useful tool for epidemiologists? *Soc Sci Med* 1987;25:905-10.
- 3 Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
- 4 McHorney CA, Ware JE, Raczek AE. The MOS 36-item short-form health survey: II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;31:247-63.
- 5 Tarlov AR, Ware JE, Greenfield S, Nelson EC, Perrin E, Zubkoff M. The medical outcomes study: an application of methods for monitoring the results of medical care. *JAMA* 1989;262:925-30.
- 6 Stewart AL, Ware JE, eds. *Measuring functioning and well-being: the medical outcomes study approach*. London: Duke University Press, 1992.
- 7 Stewart AL, Greenfield S, Hays RD, Wells K, Rogers MH, Berry SD, et al. Functional status and well-being of patients with chronic conditions: results from the medical outcomes study. *JAMA* 1989;262:907-13.
- 8 Frater A, Costain D. Any better? Outcome measures in medical audit: patients care about outcome—should doctors be measuring it? *BMJ* 1992;304:519-20.
- 9 McHorney CA, Ware JE, Rogers W, Raczek AE, Lu JFR. The validity and relative precision of MOS short- and long-form health status scales and Dartmouth COOP charts: results from the medical outcomes study. *Med Care* 1992;30(suppl):MS253-65.
- 10 Brazier JE, Harper R, Jones NMB, O' Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992;305:160-4.

- 10a Jenkinson C, Coulter A, Wright L. Short form 36 (SF 36) health survey questionnaire: normative data for adults of working age. *BMJ* 1993;306:1437-40.
- 11 Streiner GL, Norman DR. *Health measurement scales: a practical guide to their development and use*. Oxford: Oxford University Press, 1990.
- 12 Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297-334.
- 13 Kline P. *A handbook of test construction*. London: Methuen, 1986.
- 14 Nunnally JC. *Psychometric theory*. 2nd ed. New York: McGraw-Hill, 1978.
- 15 Child D. *The essentials of factor analysis*. 2nd ed. London: Cassell, 1990.
- 16 Jolliffe IT. *Principal component analysis*. New York: Springer Verlag, 1986.
- 17 Vickrey BG, Hays RD, Graber J, Rausch R, Engel J, Brook RH. A health-related quality of life instrument for patients evaluated for epilepsy surgery. *Med Care* 1992;30:299-312.
- 18 Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey: reliability and validity in a patient population. *Med Care* 1988;26:724-33.
- 19 Armitage P, Berry G. *Statistical methods in medical research*. 2nd ed. Oxford: Blackwell, 1987.
- 20 Garratt AM, Macdonald LM, Ruta DA, Russell IT, Buckingham JK, Krukowski ZH. Towards the measurement of outcome for patients with varicose veins. *Quality in Health Care* 1993;2:5-10.
- 21 Ellwood PM. Outcomes management: a technology of patient experience. *N Engl J Med* 1988;318:1549-56.
- 22 Roland MO, Coulter AM. *Hospital referrals*. Oxford: Oxford University Press, 1992.

(Accepted 19 April 1993)

## Damage to DNA in cervical epithelium related to smoking tobacco

Andrew M Simons, David H Phillips, Dulcie V Coleman

### Abstract

**Objective**—To determine whether tobacco smoking causes increased DNA modification (adducts) in human cervical epithelium.

**Design**—Comparison of DNA adducts measured by the technique of postlabelling with phosphorus-32 in normal ectocervical epithelium of smokers and non-smokers. A questionnaire on smoking habit and a urinary cotinine assay were used to identify smokers and non-smokers.

**Setting**—Cytology unit in large teaching hospital.

**Subjects**—39 women (11 current smokers, seven former smokers, and 21 who had never smoked) undergoing gynaecological treatment (colposcopy or hysterectomy). Nineteen members of staff who did not smoke as controls.

**Interventions**—Biopsy of normal ectocervical epithelium. Urine sample.

**Main outcome measures**—Measurement of DNA adducts in cervical epithelial tissue of smokers and non-smokers. Smoking habit derived from results of questionnaire and urinary cotinine:creatinine ratio. Proportion of adducts in women with abnormal and normal results of cervical smear test.

**Results**—DNA samples from smokers (identified from questionnaire) had significantly higher median proportions of DNA adducts than non-smokers (4.62 (95% confidence interval 4.04 to 7.74) v 3.47 (2.84 to 4.78) adducts/10<sup>6</sup> nucleotides; *p*=0.048). Exclusion of women whose urinary cotinine:creatinine ratio did not confirm their self reported smoking habit (smoker or non-smoker) increased this difference (4.7 (3.85 to 8.08) v 3.52 (2.32 to 4.95) adducts/10<sup>6</sup> nucleotides; *p*=0.03). Women who had abnormal results of cervical smear tests had significantly higher proportions of adducts than those with normal results (4.7 (3.90 to 8.13) v 3.47 (3.06 to 5.36) adducts/10<sup>6</sup> nucleotides; *p*=0.03).

**Conclusions**—Tobacco smoking by women leads to increased modification of DNA in cervical epithelium, suggesting biochemical evidence consistent with smoking as a cause of cervical cancer.

### Introduction

Numerous epidemiological studies have shown an association between smoking and cervical cancer.<sup>1</sup> The evidence for this association is based on population and case-control studies. Many of the studies in which other known risk factors for cervical cancer are adjusted for have shown that women who smoke have up to four times higher risk of developing cervical cancer than non-smokers.<sup>2</sup> These epidemiological studies, however, do not provide the essential molecular evidence to show whether this relation between smoking and cervical cancer is casual or causal. In 1986

the International Agency for Research on Cancer (IARC) reviewed the available epidemiological data and concluded that there was insufficient evidence at that time to include smoking as a causal agent of cervical cancer.<sup>3</sup>

Many chemical carcinogens exert their biological activity through covalent modification of DNA to form adducts. In the human respiratory tract, where the association between smoking and cancer is unequivocal,<sup>3</sup> it has been clearly shown that DNA from smokers has significantly higher proportions of adducts than that of non-smokers.<sup>4,7</sup> Also cotinine, a metabolite of nicotine, has been detected in the cervical mucus of women, showing that the cervical epithelium is exposed to components of tobacco smoke.<sup>8,9</sup>

We investigated the presence of DNA adducts in normal squamous cervical epithelium by using postlabelling with phosphorus-32 and compared the results with the smoking habits of the women.

### Subjects and methods

Women requiring a hysterectomy for benign disease or colposcopy after abnormal results of a cervical smear test were recruited into the study. They were interviewed on admission and asked to complete a questionnaire which asked whether they had ever (currently or previously) smoked. If so, then the length of time they had smoked, the number of cigarettes they smoked a day, and last time they smoked were recorded. Women who reported that they had never smoked were recorded as such. All women were asked whether they had smoked in the preceding 48 hours. Questions on passive exposure to tobacco carcinogens and general health were included. Approval from an ethical committee and informed consent were gained for participation in the study.

A urine specimen was collected at the time of interview to assay for cotinine. The pH was recorded and the sample frozen at -20°C until analysis. Urine samples were also collected from 19 medical staff (known non-smokers) to act as controls.

A punch biopsy specimen of normal ectocervical epithelium (acetowhite negative areas) was obtained from each woman undergoing colposcopy. A biopsy specimen of ectocervical epithelial tissue (about 6 mm × 10 mm) was excised from the anterior lip of the cervix of women who had a hysterectomy. The samples were frozen immediately and stored at -20°C until analysis.

**DNA extraction**—DNA was isolated from thawed biopsy specimens essentially as described previously.<sup>10</sup> The DNA yield from each sample was determined spectrophotometrically.

**<sup>32</sup>P postlabelling**—Samples of 4 µg of DNA were

Cytology and Cytogenetics Unit, St Mary's Hospital Medical School, Faculty of Medicine, Imperial College, London W2 1PG

Andrew M Simons, research fellow  
Dulcie V Coleman, professor

Haddow Laboratories, Institute of Cancer Research, Surrey SM2 5NG

David H Phillips, senior scientist

Correspondence to: Professor Coleman.

*BMJ* 1993;306:1444-8