Are postal questionnaire surveys of reported activity valid? An exploration using general practitioner management of hypertension in older people

MARTIN ECCLES

GARY A FORD

SANDRA DUGGAN

NICK STEEN

SUMMARY

Background. *Postal questionnaire surveys are commonly used in general practice and often ask about self-reported activity. The validity of this approach is unknown.*

Aim. *To explore the criterion validity of questions asking about self-reported activity in a self-completion questionnaire.*

Method. *A comparison was made between (a) the self-reported actions of all general practitioner (GP) principals in 51 general practices randomly selected within the nine family health services authorities of the former northern regional health authority, and (b) the contents of the medical records (case notes and computerized records) of patients classified as hypertensive from a 1 in 7 random sample of all patients registered in these practices and aged between 65 and 80. Data were gathered from the GPs by self-completion postal questionnaires. Six comparisons were made for two groups of items: first, target and achieved blood pressure; secondly, patient's weight, smoking status, alcohol consumption, exercise and salt intake. The frequency with which the data items were recorded in patient records was compared with the GPs' self-reported frequency of performing the actions.*

Results. *No relationship was found between achieved blood pressure and stated target levels. For each of the other actions, more than half of the responders reported that they usually or always performed the activity. For four of these (smoking, weight, alcohol and exercise), a significant association was noted, but the size of this varied considerably.*

Conclusions. *There is a variable relationship between what responders report that they do in self-completion questionnaires, and what they actually do as judged by the contents of their patients' medical records. In the absence of prior knowledge of the validity of questions on reported activity, or of concurrent attempts to establish their validity, the questions should not be asked.*

Keywords: questionnaire; criterion validity.

M Eccles MD, FRCP, FRCGP, professor of clinical effectiveness; S Duggan, MSc, research associate; and N Steen, PhD, research associate, Centre for Health Services Research, University of Newcastle Upon Tyne. G A Ford, FRCP, senior lecturer, Department of Medicine and Pharmacological Sciences, University of Newcastle-upon-Tyne.

Submitted: 17 December 1997; final acceptance: 19 August 1998.

Introduction

SURVEYS are commonly used as a method of gathering opinion. When the aim is to gather information on responders' attitudes, knowledge and beliefs, few alternative methods exist. However, surveys are also used to gather additional information on the reported actions of the responder.[1-6] While methods of establishing the validity of opinion surveys are well established,[7] the validity of questions on reported activity is not addressed by the same methods. For such questions, validity is addressed by comparison against an external data source, such as medical records or direct observation. The validity so obtained is called criterion validity. It is seldom reported and the assumption seems to be that reported results are valid. However, there are grounds for assuming that this is not the case.

If the intention is to know about the activities of health care professionals, relying on their perceptions may be misleading. Montano *et al*[8] compared physician self-report with chart audit and patient survey; data from the last two methods were highly correlated, but the relationship between physician self-report and either of the other two methods was much weaker. As one part of a larger study of the management of hypertension in older people,[9,10] the authors compared the self-reported actions of GPs with the medical records of a random sample of their older patients with hypertension.

Methods

All GP principals in 51 general practices randomly selected within the nine family health services authorities of the former Northern Regional Health Authority were asked to complete a questionnaire about their management of older patients with hypertension. The questionnaire asked about specific areas of their management of hypertension in older people, their attitudes towards the management of hypertension in this age group,[10] factors that they perceived as influencing their management, and demographic details. The questionnaires were delivered to each practice by the study data collector.

The medical records (case notes and computerized records) of a 1-in-7 random sample of all patients between the ages of 65 and 80 were examined by a nursing-qualified data collector. The following data were collected for the preceding five years from patients' notes and for the whole period covered by their computerized records:

- patient demographics,
- levels of blood pressure,
- whether or not patients were recorded as being hypertensive,
- chronic illnesses relevant to hypertension or its treatment, and
- drug treatments.

Using a validated algorithm, hypertensive patients were identified.[9] The analysis was restricted to these patients. From two questions, six comparisons were made between the reported

actions of the doctor and the level of recording of the action within the medical records. For the question 'What is your target blood pressure in an otherwise well 75-year-old patient with sustained hypertension who is tolerating antihypertensive therapy?' the comparison was between the mean reported target systolic blood pressure of all the responding GPs in the practice and the mean systolic blood pressure of treated hypertensive patients in the practice. The second question, which was answered on a five-point Likert scale (never, rarely, sometimes, usually, always), was: 'In the assessment of newly diagnosed older hypertensives, how often would you or your practice staff enquire about the following: weight, smoking status, exercise, salt intake, alcohol intake.' Here, the comparison was between the responses of the doctors and the recording of the items in the patients' records.

Any of the five variables was deemed to have been enquired about if a patient's records contained an explicit note such as 'weight discussed with patient' or 'advised to stop smoking'. In addition, weight was deemed to have been discussed if the patient's weight or body mass index was recorded; similarly, any explicit record of the number of cigarettes the patient smoked, the number of units of alcohol taken, the amount of salt ingested, or the amount of exercise taken indicated that the issue had been addressed.

Patients were registered with a general practice, but it was not possible to identify an individual GP responsible for the management of their hypertension. Standard statistics such as Cohen's Kappa or Kendall's coefficient of concordance[11] were therefore not applicable. Instead, multilevel modelling[12] was used to test the hypothesis that as the proportion of doctors in a practice who reported a particular activity increased, so did the proportion of patients for whom that activity was recorded in their records. In these analyses, the dependent variable was measured at the patient level but the explanatory variable was a practice characteristic. It was necessary to use multilevel modelling to allow for the lack of independence between observations corresponding to different patients within a particular practice.

Five of the dependent variables (weight, smoking status, exercise, salt intake and alcohol intake) were binary — either the item was recorded in the patient's records or it was not. The model adopted was the multilevel binary response model implemented in MLn (a standard logistic model was used to model variation between patients, and variation in recording behaviour between practices was included as a random effect with an assumed normal error distribution). Practices in which all GPs reported undertaking the activity were then compared with all the other practices (for recorded exercise, the comparison was between practices where at least 75% of doctors reported about exercise, and practices where this was not the case). This was done by fitting a variable that took the value 0 or 1, as appropriate, for all patients within a particular practice. The parameter estimate corresponding to this variable and its associated standard error was used to generate an interval estimate for the difference between the two types of practice in the form of an odds ratio.

The remaining dependent variable, achieved blood pressure, was the mean of the most recent systolic blood pressures (up to a maximum of three) in the patient's notes. Variations in achieved blood pressure between patients and between practices were included as random effects. The mean stated target blood pressure was included as an explanatory variable (this was the mean of individual responses of GPs within a practice and took the same value for all patients within a practice). The parameter estimate corresponding to this variable and its associated standard error were used to derive an interval estimate for the strength of the association between achieved blood pressure and stated target blood pressure.

Results

Five single-handed GPs did not return questionnaires, so the details of their patients were excluded from the analysis. Thus the analysis was based on the 1672 hypertensive patients in 46 practices. A total of 194 questionnaires were distributed and 158 returned, an overall response rate of 81%. Seventeen responders did not answer the second question, giving a response rate of 73% for that section of the analysis. The mean (SD) proportion of GPs per practice who completed questionnaires was 0.85 (0.2).

The relationship between stated target and achieved systolic blood pressure is shown in Figure 1. This was investigated using multilevel modelling. The model fitted assumed that there were differences between practices (modelled as a random effect) and that there was random variation in systolic blood pressure between patients within practices. The model suggested that differences between practices were significant; the intraclass correlation coefficient (a measure of the extent to which blood pressures of patients within a practice are correlated) was 0.04 (95% CI = 0.01 to 0.07). The mean stated target blood pressure was then entered as an explanatory variable. It explained virtually none of the practice level variation, and the regression coefficient was 0.01 (95% CI = –0.17 to 0.19). This analysis confirms the impression given by Figure 1 that there is no relationship between stated target and achieved blood pressure.

The overall responses to the second question are shown in Table 1. Although there is a spread of responses, for each action more than half of the responders reported that they usually or always performed it. The results of the modelling are shown in Table 2; the reported odds ratios are the likelihood of the activity being recorded in a patient's records if all the doctors in the practice report that they always enquire about it compared with the likelihood of it being reported in any other circumstance. For the four comparisons, the associations are all statistically significant; activities were more likely to be recorded for patients in practices

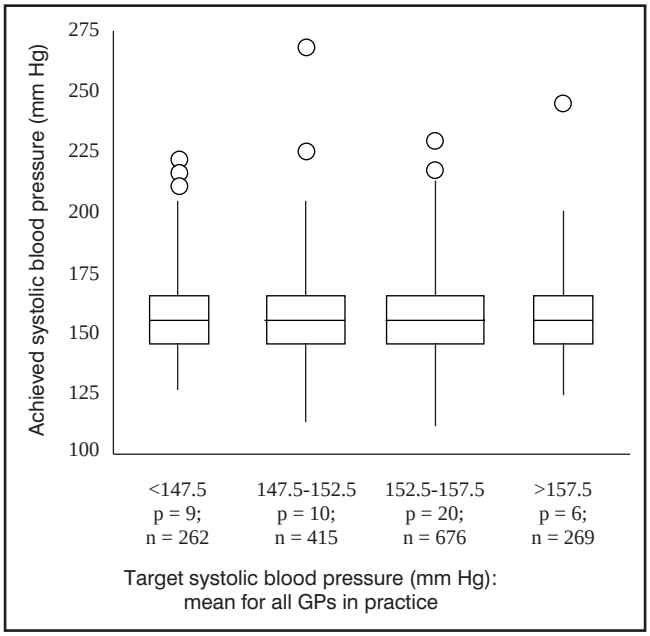

Figure 1. Practice mean target systolic blood pressures and achieved mean systolic blood pressures (p = number of practices corresponding to each plot; n = number of patients). The lines through the middle of the boxes indicate the median blood pressure, and the upper and lower boundaries of the boxes represent the 25th and 75th percentiles; thus 50% of patients have a blood pressure that falls within the boundaries of the relevant box.

where all GPs reported carrying out that activity. In general, the differences were largest for those variables where the overall level of recording was smallest (although the confidence intervals tended to be widest for these variables). Only 10 patients had their salt intake recorded in their records; as this was clearly unrelated to the reported level of enquiry for this variable formal statistical analysis was not warranted.

Discussion

The results of this study show a variable relationship between what the responders to a self-completion questionnaire report they do and what they actually do, as judged by the contents of their patients' records. This relationship ranges from good agreement for smoking status to no agreement at all for salt intake and blood pressure control.

The results suggest that a number of factors may affect the criterion validity of questions of self-reported behaviour. These include the nature of the activity concerned and the frequency with which that activity is performed. For reported target and achieved blood pressure (one of the comparisons in which no relationship was demonstrated), the situation is complex. Although a responder may state a single target value for the control of blood pressure, patients start from differing pre-treatment levels and there are many factors that may influence whether or not a patient's blood pressure falls; these include response to treatment and compliance. The responder's target value may be accurate, but the achieved value is influenced by multiple factors that cannot be understood from a comparison against a value derived from a cross-sectional study. In such cases, validation could be more properly derived from a comparison against longitudinal data, which allows a clearer understanding of a complex and continuing process of care.

The other comparisons were of much simpler actions but were influenced by the frequency with which they were recorded as being performed. As the frequency drops, the odds ratio rises but the confidence interval widens. These results are likely to be an underestimate of the association because a proportion of the patients would have been newly diagnosed before the period covered by the notes-based search performed in this study. Data recorded in these patient's notes more than five years ago and not appearing in their computer records would not have been abstracted. However, there is no reason to assume that this would differentially affect the five activities. Thus, in the case of simple actions, where a comparison is more reasonable, there is no constant relationship between self-reported and recorded activity. This means that although self-reported data might be usable for a relative comparison within a single variable, it is an unreliable comparator between activities. Report of 'always performing' an activity at two points in time may relate to a constant level of activity. However, report of 'always performing' certainly relates to differing, and thus non-comparable, levels of activity.

However, in interpreting these results a number of limitations of this study need to be explored. The comparison to assess criterion validity relied upon the contents of the patients' medical records. Although anecdotally the content of general practice records is perceived to be at best variable and at worst poor, it is difficult to find systematic attempts to quantify their reliability. Restricted space and management by exception result in only abnormal findings being recorded;[13] providing more space within a structured record results in a much fuller account of care, but such record keeping is seen as time consuming.[14] Thus, if an activity is not recorded in a patient's medical record, it is possible that the activity has been performed but not recorded. However, Donabedian[15] has argued that good doctors keep good records. Furthermore, if medical records are acceptable in a court of law, as an account of the care provided, they should also be acceptable as a means of judging quality of care.[16]

The validity of data in patient records is supported, at least in secondary care, by two studies. Montano *et al*[8] demonstrated a strong correlation between clinicians' activities as assessed by chart audit and by patient survey. Kahn *et al*[17] demonstrated that, for four out of five assessed conditions, better documentation of the process of care was associated with improved outcome in the form of lower mortality rates 30 days after admission to hospital. In primary care, Wilson and McDonald[18] demonstrated that the medical record underestimated the recording of advice about smoking and alcohol in the context of health promotion. However, the sensitivity of patient records was the same for both items. Thus, using patients' records as a pragmatic indicator of actions performed is reasonable.

An ideal assessment of the validity of postal questionnaires on self-reported activity would examine different survey methods, and would compare what GPs say they do with both (a) data from their patients' medical records, and (b) direct observations using methods such as simulated patients, or video recordings of the activities in question. Such a comparison would be time consuming and expensive to conduct. Perhaps a minimum requirement in any questionnaire-based survey would be to state what has been done to determine the validity of the instrument, so that readers can better assess the value of its results. Without prior

Table 1. Percentage responses from 141 GPs to the question: 'In the assessment of newly diagnosed older hypertensives, how often would you or your practice staff enquire about the following: weight, smoking status, exercise, salt intake, alcohol intake?'

Variable	Never	Rarely	Sometimes	Usually	Always
Smoking	0	0.7	2.1	20.6	76.6
Weight	0	0	6.4	41.4	52.1
Alcohol	0.7	2.8	12.8	38.3	45.4
Exercise	0	7.1	22.0	39.0	31.9
Salt intake	5	13.5	31.2	27.7	22.7

Table 2. Percentage of patients' records containing entries on smoking, weight, alcohol, and exercise for three categories of practice.

	Percentage of records mentioning variable			
Variable	Overall	Practices where all GPs reported always enquiring about it	Practices where GPs reported any other frequency of recording	Odds ratio (95% CI)
Smoking	74.8	79.6	71.4	1.73 (1.09 to 2.76)
Weight	60.9	72.3	59.1	2.02 (1.16 to 3.55)
Alcohol	59.0	80.1	56.6	2.94 (1.43 to 6.04)
Exercise	13.9	26.9[a]	12.7	3.54 (1.23 to 10.1)

[a]For exercise, the percentage refers to practices in which at least 75% of GPs reported always enquiring about the variable.

knowledge of the validity of questions on reported activity, or without concurrent attempts to establish their validity, the questions should not be asked. Given the increasing use of survey methods and the perceived load on responders,[19] researchers and funders of research should ensure that instruments are as valid as possible.

References

1. Duran-Tauleria E, Rona RJ, Chinn S, Burney P. Influence of ethnic group on asthma treatment in children in 1990-1: national cross sectional study. *BMJ* 1996; **313:** 148-152.
2. Ronsmans C, Islam T, Bennish ML. Medical practitioners' knowledge of dysentry treatment in Bangladesh. *BMJ* 1996; **313:** 205-206.
3. Mason D, Kerry S, Oakeshott P. Postal survey of the management of cervical *Chlamydia trachomatis* infection in English and Welsh general practices. *BMJ* 1996; **313:** 1193-1194.
4. Strang J, Sheridan J, Barber N. Prescribing injectable and oral methadone to opiate addicts: results from the 1995 national postal survey of community pharmacies in England and Wales. *BMJ* 1996; **313:** 270-272.
5. Sheridan J, Strang J, Barber N, Glanz A. Role of community pharmacies in relation to HIV prevention and drug misuse: findings from the 1995 national survey in England and Wales. *BMJ* 1996; **313:** 272-274.
6. While D, Kelly S, Huang W, Charlton A. Cigarette advertising and onset of smoking in children: questionnaire survey. *BMJ* 1996; **313:** 398-399.
7. Moser C, Kalton G. *Survey methods in social investigation*. 2nd edition. Aldershot: Gower, 1971.
8. Montano DE, Phillips WR. Cancer screening by primary care physicians: a comparison of rates obtained from physician self-report, patient survey, and chart audit. *Am J Pub Health* 1995; **85:** 795-800.
9. Duggan S, Aylett MJ, Eccles M, Ford GA. Defining hypertension in older people from primary care case notes review. *J Hum Hyp* 1997; **11:** 193-199.
10. Duggan S, Ford GA, Eccles M. Doctors' attitudes towards the detection and treatment of hypertension in older people. *J Hum Hyp* 1997; **11:** 271-276.
11. Seigel S, Castellan NC. *Non-parametric statistics for the behavioural sciences*. 2nd edition. New York: McGraw-Hill, 1988.
12. Rasbash, J, Woodhouse G. *MLn command reference version 1.0*. Multilevel Models Project, Institute of Education, University of London, 1995.
13. Benjamin B. *Medical records*. 2nd edition. London: Heinemann, 1977.
14. Centre for Health Services Research. *North of England study of standards and performance in general practice. Report 42: volume III. The effects of setting and implementing clinical standards.* Centre for Health Services Research, Newcastle upon Tyne, 1991.
15. Donabedian A. *A guide to medical care administration. Volume II: Medical care appraisal*. New York: American Public Health Association, 1969.
16. Osborne CE. Interdiagnosis relationships of physician recording in ambulatory child health care. *Med Care* 1977, **15:** 465-474.
17. Kahn K, Rogers WH, Rubenstein LV, *et al*. Measuring quality of care with explicit process criteria before and after implementation of the DRG-based prospective payment system. *JAMA* 1990; **264:** 1969-1973.
18. Wilson A, McDonald P. Comparison of patient questionnaire, medical record, and audio tape in assessment of health promotion in general practice consultations. *BMJ* 1994; **309:** 1483-1485.
19. McAvoy BR, Kaner EFS. General practice postal surveys: a questionnaire too far? *BMJ* 1996; **313:** 732-733.

Acknowledgements

This study was jointly funded by the NHS R&D Cardiovascular Disease and Stroke Programme and the R&D division of the former Northern Region Health Authority. We are grateful to all the GPs who participated in the study and to Samantha Jones for collecting the data from the medical records.

Address for correspondence

Professor M Eccles, Centre for Health Services Research, University of Newcastle Upon Tyne, 21 Claremont Place, Newcastle-upon-Tyne NE2 4AA. Email: martin.eccles@ncl.ac.uk