

Evidence of self-report bias in assessing adherence to guidelines

ALYCE S. ADAMS¹, STEPHEN B. SOUMERAI¹, JONATHAN LOMAS² AND DENNIS ROSS-DEGNAN¹

¹Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, Boston and the

²Canadian Health Services Research Foundation, Ottawa, Canada

Abstract

Objective. To assess trends in the use of self-report measures in research on adherence to practice guidelines since 1980, and to determine the impact of response bias on the validity of self-reports as measures of quality of care.

Methods. We conducted a MEDLINE search using defined search terms for the period 1980 to 1996. Included studies evaluated the adherence of clinicians to practice guidelines, official policies, or other evidence-based recommendations. Among studies containing both self-report (e.g. interviews) and objective measures of adherence (e.g. medical records), we compared self-reported and objective adherence rates (measured as per cent adherence). Evidence of response bias was defined as self-reported adherence significantly exceeding the objective measure at the 5% level.

Results. We identified 326 studies of guideline adherence. The use of self-report measures of adherence increased from 18% of studies in 1980 to 41% of studies in 1985. Of the 10 studies that used both self-report and objective measures, eight supported the existence of response bias in all self-reported measures. In 87% of 37 comparisons, self-reported adherence rates exceeded the objective rates, resulting in a median over-estimation of adherence of 27% (absolute difference).

Conclusions. Although self-reports may provide information regarding clinicians' knowledge of guideline recommendations, they are subject to bias and should not be used as the sole measure of guideline adherence.

Keywords: clinical competence, physician practice patterns, practice guidelines, process assessment, quality assurance, quality of care measurement

Continuing evidence of medical practice variations and gaps in the quality of care have spurred the rapid development of practice guidelines in most areas of clinical practice. There has been a commensurate increase in studies examining clinician adherence to these guidelines as process measures of quality of care. However, the validity of this literature may be compromised by an over-reliance on self-report measures of guideline adherence because of possible response biases in self-reports [1,2].

This study sought to answer the following questions: (i) has the use of self-report measures of guideline adherence in published research increased since 1980?; (ii) are these measures being used to assess quality of care?; (iii) do comparisons with more objective (unobtrusive) measures confirm the existence of substantial over-estimation of adherence in self-report measures?

Methods

Literature search

To identify literature pertaining to clinician adherence to guidelines, we conducted a search of the medical literature on MEDLINE, written in English, using the terms: practice guidelines, physician practice patterns, continuing medical education, medical audit, utilization review, quality assurance, quality of health care, drug utilization review, clinical competence, nursing audit, outcomes/process assessment, peer review, evaluation studies, official policy, hospital standards, standards and health services research for the period 1980–1996. We also searched the literature for work by key authors in the field of guidelines research.

Address correspondence to Stephen B. Soumerai, Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, 126 Brookline Avenue, Suite 200, Boston, MA 02215, USA. Tel: +1 617 421 6863. Fax: +1 617 421 2763. E-mail: stephen_soumerai@hms.harvard.edu

Inclusion criteria

In order to establish trends in the use of objective and self-report measures of adherence over time, we included only studies that: (i) evaluated the adherence of clinicians to practice guidelines, official policies, or other evidence-based recommendations; (ii) specified the guidelines being utilized; (iii) identified the population to which they apply; and (iv) reported the recommended periodicity for each preventive service and the indications for each therapeutic service examined. We categorized the studies that were identified according to the type of method used to measure guideline adherence. Self-report measures included self-administered questionnaires and face-to-face interviews. Some questionnaires included patient vignettes, followed by questions ascertaining clinicians' preventive or therapeutic decision making in response to the case described. The objective measures included review of medical records, discharge data, prescriptions, claims data, or observation of actual practice. In two studies, unobtrusive observation of actual practice included the use of simulated cases in which anonymous trained actors presented themselves to clinicians as new patients with a standardized set of symptoms and/or risk factors specified by the researcher. At the end of the visit, the actor recorded the preventive or therapeutic services or actions taken by the clinician.

Exclusion criteria

To reduce the potential for measurement error, studies that used measures with questionable objectivity or those for which adherence rates were unclear were excluded. For example, studies that compared physician self-reports to surveys of patients were excluded because of the susceptibility of patient surveys to a variety of observer biases. In addition, studies using videotaped encounters were excluded because the awareness of being observed may cause clinicians to diverge from their usual behaviours [3]. Measures that could not differentiate between preventive (screening) or diagnostic uses of the tests (e.g. sigmoidoscopy) based on the information provided were also excluded.

Analysis

Adherence to guidelines was measured by dividing the percentage of the patient population (or expected number of patients) receiving the service in question, obtained using either self-report or objective methods, by the percentage (or expected number) who should have received the service according to the guidelines. Study findings supported the presence of response bias if the rate of self-reported adherence with recommended practice exceeded the objective adherence rate. We determined statistical significance by calculating the χ statistic for differences in proportions for a one-tailed test at the 0.05 level.

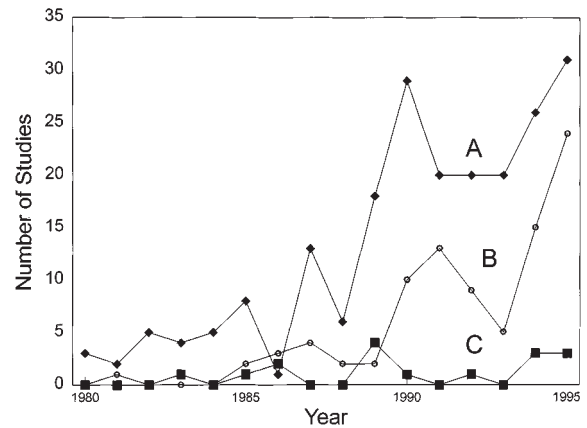


Figure 1 Trends in the use of objective and self-report measures in studies of guideline adherence. A, Objective measures only; B, self-report only; C, both self-report and objective measures.

Results

Trends in the use of self-report measures

Using the search criteria defined above, we identified 326 studies published between January, 1980 and June, 1996. The majority of studies examined physician adherence to guidelines, although some authors studied the practice patterns of nurses, dentists, and pharmacists. There was a strong upward trend in the number of such studies published since the late 1980s (Figure 1). The majority of these studies used only measures of adherence that were obtained objectively. However, the gap between the use of only objectively obtained measures and only self-reported measures has decreased considerably over time. By 1995, 41% of all guideline adherence studies relied exclusively on self-reports.

Between 1992 and 1996, virtually all self-report studies were primarily intended to assess actual practice behaviours. Overall, 65% of self-report studies examined compliance with guidelines for treatment or diagnostic services and 29% with guidelines for preventive care. (The focus of 6% of these studies could not be classified.) Approximately 14% of studies using self-report measures estimated the effect of educational or training programmes on practice patterns.

Concurrent use of self-report and objective measures

Twenty studies used both objective and self-report measures; 10 of these were excluded from the analysis because of our inability to identify comparable objective and self-report measures based on the information provided [4–13]. Of the remaining 10 studies [14–23] four examined adherence to guidelines for preventive care [14,15,17,20], one investigated rates of Caesarean sections [18], one observed preventive services for symptomatic populations at risk for sexually transmitted diseases [21] and four examined adherence to drug treatment guidelines for acute myocardial infarction, childhood diarrhoea, and anxiety [16,19,22,23] (see Table 1

Table 1 Summary of findings of existence of bias in self-reports

Author (year) [reference]	Service provided	Practice standard	Self-report method	Objective method	Self-report	Objective	NS ¹
McPhee <i>et al.</i> (1986) [2]	Cancer screening	American Cancer Society	Face-to-face interview	Chart review	6	0	0
Weingarten <i>et al.</i> (1995) [20]	Cancer screening and influenza vaccine	U.S. Preventive Task Force	Self-administered questionnaire	Chart review	1	0	2
Headrick <i>et al.</i> (1992) [17]	Hypercholesterolaemia screening	National cholesterol education programme	Self-administered questionnaire	Chart review	1	0	0
Battista (1983) [14]	Smoking cessation recommendations	Canadian Task Force	Face-to-face interview	Billing data	1	0	0
Rabin <i>et al.</i> (1994) [21]	Sexually transmitted disease prevention	U.S. Preventive Task Force	Self-administered questionnaire	Simulated patient	5	3	0
Lomas <i>et al.</i> (1989) [18]	Caesarean section	Canadian Medical Association	Self-administered questionnaire	Discharge data	1	0	1
McLaughlin <i>et al.</i> (1996) ² [16]	Drug treatment: acute myocardial infarction	AHA ³ /American College of Cardiology	Self-administered questionnaire	Prescription data	4	0	0
Yeo <i>et al.</i> (1994) [22]	Drug treatment: anxiety and insomnia	RACGP ⁴ intervention	Face-to-face interview	Prescription pads	2	0	0
Ross-Degnan <i>et al.</i> (1996) [23]	Drug treatment: acute diarrhoea	World Health Organisation	Self-administered questionnaire	Simulated patient/drug sales	4	0	0
Nizami <i>et al.</i> (1995) [19]	Drug treatment: acute diarrhoea	World Health Organisation	Face-to-face interview	Trained observers	5	1	1

¹ NS, not significant.² Includes unpublished data from this study provided by S. B. Soumerai.³ AHA, American Heart Association.⁴ RACGP, Royal Australian College of General Practitioners.

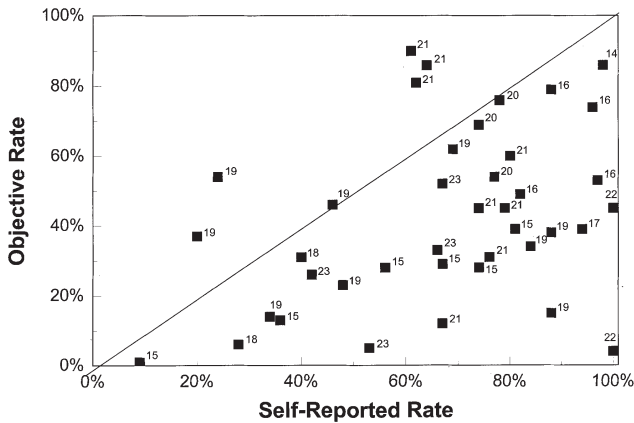


Figure 2 Self-reported versus objective adherence. Observations falling below the diagonal provide evidence of self-report bias; numbers to the right of each observation are the reference numbers of the relevant studies.

for a description of the studies included). Three studies were conducted in developing nations, two in Canada, one in Australia, and the remaining six in the USA. All but two studies examined the behaviour of physicians. The exceptions were a study of preventive care practices of both nurse practitioners and physicians [15], and a study of drug prescribing by pharmacists [23].

Of the self-report measures employed in the 10 identified studies, four studies used face-to-face interviews, six used self-administered questionnaires, and two used patient vignettes. Of the objective measures, five used medical records or administrative data as the primary data source and two used drug prescriptions or actual drugs provided (Table 1).

Extent of over-estimation

The majority of studies support the existence of a bias in self-reports (Table 1 and Figure 2). On average, clinicians tended to over-estimate their adherence to recommended norms by a median absolute difference of 27%. Overall, 87% of the 37 self-reported measures examined in these studies over-estimated actual adherence to guidelines, and 32 (88%) of these differences were statistically significant.

Some of the preventive studies included in this analysis examined the same services. For four of the services (mammography, breast exam, rectal exam, test for occult blood), the results consistently supported the existence of over-estimation. The two services for which self-report did not always exceed objective compliance were Papanicolaou smear and sigmoidoscopy.

There are five differences (out of 37) depicted in Figure 2 that do not demonstrate over-estimation. Three were from the same study, and involved especially sensitive topics [21]. The other two differences are also interesting in that the respondents, general practitioners, were considered by the authors to be unaware of the guidelines and would therefore be less susceptible to biased responses [19].

Discussion

Our analysis of published research on clinician adherence to practice guidelines from 1980 to 1996 indicates a substantial increase in the use of self-report measures over time. Moreover, 83% of these self-reports were intended to estimate actual adherence to guidelines. The extent of bias is substantial; the median absolute magnitude of over-estimation was 27%. Thus, the increasing reliance on self-reports as a measure of quality of care appears to produce gross over-estimation of performance. It is especially disconcerting that the magnitude of bias is greater than the degree of improvement observed after many guideline implementation interventions [24].

Two plausible types of bias that may explain this over-estimation are social desirability and interviewer bias. Social desirability bias occurs when an individual does not adhere to a social norm, but reports the socially desirable behaviour when questioned [25,26]. Failure to comply with a social norm suggests the existence of: (i) forces that make change from current practice equally or more discomforting than non-adherence; (ii) rewards for not changing behaviour; or (iii) other barriers to change [27]. We hypothesize that the process of guideline dissemination exposes clinicians to social pressures that may promote socially desirable responses that do not reflect actual practice when clinicians face impediments to adherence to the guidelines.

The reviewed studies provided examples of the forces that lead to socially desirable responses. When clinicians were asked why they failed to actually comply with guidelines, they identified specific training or knowledge deficits as well as barriers to adherence that were beyond their control, such as perceived threats of malpractice litigation, inadequate skills, and economic and socio-economic incentives to perform non-recommended services [18,19,23]. In addition, they cited systemic or extra-systemic factors that precluded compliance, such as the unavailability of equipment or supplies, and administrative or logistic barriers [15]. As indicated by the Nizami study, it is possible that those clinicians who have the greatest degree of exposure to guidelines are more susceptible to this type of bias [19]. Therefore, our findings would be further strengthened by research that showed a definitive link between social desirability bias and the physician's awareness of the guidelines.

Interviewer bias is closely related to social desirability bias, except that the respondent provides the response that he or she believes the interviewer wants to hear [26]. We would expect this form of bias to be greatest when the interviewer is a respected colleague or leader in the field. The studies in this review do not contain adequate information to allow distinctions between interviewer types. Due to its potential to invalidate findings based on self-report, interviewer bias in studies of guideline adherence is a topic that deserves further study.

Alternative explanations for over-estimation of adherence include memory errors or the unreliability of objective measures. There are two types of memory errors in surveys to recall events: forgetting, which results in under-reporting; and

telescoping, in which the respondent remembers an event as occurring more recently than it did [28]. Telescoping can lead to over-estimation, and its impact in self-report measures remains to be critically examined. Errors in estimation would be expected to be random and should result in both under- and over-estimation. Only some forms of systematic bias would cause the pattern of consistent over-estimation seen in the studies examined.

Objective measures can also be unreliable: it is possible that practitioners may perform some screening tests (e.g. breast exams) and fail to record them in the medical records [15]. This under-reporting would widen the gap between objective and self-report measures, leading us to conclude that the bias in self-reports is worse than it really is. However, our inclusion of a variety of objective measures and services suggests that such errors, however plausible, do not explain fully the magnitude and consistency of the differences presented here.

Several other limitations of our review and analysis merit discussion. The first is the small number of reviewed studies. Although only 10 studies met the inclusion criteria, these included both preventive and treatment studies, and 37 clinical practices measured by both methods. Furthermore, these studies produced consistent findings of over-estimation.

Our study includes a number of different measurement methods with varying degrees of subjectivity. The aggregation of measures into two broad categories may imply that we attributed the same degree of objectivity to billing data, chart review, and the simulated case method. Likewise, our self-report category treated self-administered questionnaires and face-to-face interviews as equally susceptible to bias. More studies are needed to determine the generalizability of social desirability and interviewer bias across different measurement methods and different guidelines.

In addition, some concerns may arise regarding the use of data from developing countries. Two of the 10 studies included were conducted in developing countries, namely Kenya, Indonesia and Pakistan. However, the consistency of the evidence for over-estimation among the remaining eight studies demonstrates that the evidence from developing nations does not drive our results. In fact, the study conducted in Pakistan was one of the two studies providing evidence against social desirability bias (see Table 1) [19].

In conclusion, our finding of persistent over-estimation provides a warning for journal editors. Given the increasing use of self-reports in guideline compliance studies, it is perhaps time for journal editors to send out a clear message that use of self-report as the sole measure of quality of care will no longer be considered scientifically defensible. Just as the low-response-rate survey is now considered unpublishable in reputable journals, so too should be studies that measure guideline adherence solely by self-report.

Acknowledgements

Supported by the U.S. Department of Health and Human Services, Agency for Health Care Policy and Research (Grant

No: 5R18HS07357) and the International Network for Rational Use of Drugs.

References

1. Jones TV, Gerrity MS, Earp J. Written case simulations: do they predict physicians' behavior. *J Clin Epidemiol* 1990; **43**: 805–815.
2. McPhee SJ, Bird JA. Implementation of cancer prevention guidelines in clinical practice. *J Gen Intern Med* 1990; **5**: S116–S122.
3. Adair JG. The Hawthorne effect: A reconsideration of the methodological artifact. *J Appl Psych* 1984; **69**: 334–345.
4. Carey TS, Garrett J, for the North Carolina Back Pain Project. Patterns of ordering diagnostic tests for patients with acute low back pain. *Ann Intern Med* 1995; **125**: 807–814.
5. Curry L, Purkis IE. Validity of self-reports of behavior changes by participants after a CME course. *J Med Educ* 1986; **61**: 579–584.
6. 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 Public Health* 1995; **85**: 795–800.
7. Morrell DC, Roland MO. Analysis of referral behavior: responses to simulated case histories may not reflect real clinical behavior. *Br J Gen Pract* 1990; **40**: 182–185.
8. Risucci DA, Tortolani AJ, Ward RJ. Ratings of surgical residents by self, supervisors and peers. *Surg, Gynecol and Obstet* 1989; **169**: 519–526.
9. Holmes MM, Rovner DR, Rothert ML *et al*. Methods of analyzing physician practice patterns in hypertension. *Med Care* 1989; **27**: 59–68.
10. Wenrich MD, Paauw DS, Carline JD *et al*. Do primary care physician screen patients about alcohol intake using the CAGE questions? *J Gen Intern Med* 1995; **10**: 631–634.
11. Woo B, Woo B, Cook EF *et al*. Screening procedures in the asymptomatic adult. *J Am Med Assoc* 1985; **254**: 1480–1484.
12. McDonald CJ, Hui SL, Smith DM *et al*. Reminders to physicians from an introspective computer medical record. *Ann Intern Med* 1984; **100**: 130–138.
13. McDonnell PJ, Nobe J, Gauderman WJ *et al*. Community care of corneal ulcers. *Am J Ophthalmol* 1992; **114**: 531–538.
14. Battista RN. Adult cancer prevention in primary care: patterns of practice in Quebec. *Am J Public Health* 1983; **73**: 1036–1039.
15. McPhee SJ, Richard RJ, Solkowitz SN. Performance of cancer screening in a university general internal medicine practice: comparison with the 1980 American Cancer Society guidelines. *J Gen Intern Med* 1986; **1**: 275–281.
16. McLaughlin TJ, Soumerai SB, Willison DJ *et al*. Adherence to national guidelines for drug treatment of suspected acute myocardial infarction. *Arch Intern Med* 1996; **156**: 799–805.
17. Headrick LA, Speroff T, Pelecanos HI, Cebul RD. Efforts to improve compliance with the National Cholesterol Education

- Program guidelines: results of a randomized controlled trial. *Arch Intern Med* 1992; **152**: 2490–2496.
18. Lomas J, Anderson GM, Domnick-Pierre K *et al.* Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N Engl J Med* 1989; **321**: 1306–1311.
 19. Nizami SQ, Kahn IA, Bhutta ZA. Difference in self-reported and observed prescribing practice of general practitioners and pediatricians for acute watery diarrhoea in children of Karachi, Pakistan. *J Diarrhoeal Dis Res* 1995; **13**: 29–32.
 20. Weingarten S, Stone E, Hayward R *et al.* The adoption of preventive care practice guidelines by primary care physicians: do actions match intentions? *J Gen Intern Med* 1995; **10**: 138–144.
 21. Rabin DL, Boekeloo BO, Marx ES *et al.* Improving office-based physicians' prevention practices for sexually transmitted diseases. *Ann Intern Med* 1994; **121**: 513–519.
 22. Yeo GT, De Burgh SPH, Letton T *et al.* Educational visiting and hypnosedative prescribing in general practice. *Fam Pract.* 1994; **11**: 57–61.
 23. Ross-Degnan D, Soumerai SB, Goel PK *et al.* The impact of face-to-face educational outreach on diarrhoea treatment in pharmacies. *Health Policy Planning.* 1996; **11**: 308–318.
 24. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993; **342**: 1317–1322.
 25. Plous S. *The Psychology of Judgement and Decision Making.* New York: McGraw Hill, Inc., 1993: pp. 191–204.
 26. Warwick DP, Lininger CA. *The Sample Survey: Theory and Practice.* New York: McGraw-Hill, Inc., 1975: pp. 201–202.
 27. Festinger L. *The Theory of Cognitive Dissonance.* Evanston, IL: Row, Peterson and Co., 1957.
 28. Bradburn NM. Response effects. In Rossi PH, Wright JD, Anderson AB, eds, *Handbook on Survey Research.* San Diego, CA: Academic Press, Inc., 1983: pp. 308–310.

Accepted for publication 2 December 1998