THE EFFECT OF REGULARITY ON THE ACCURACY OF REPORTING OF MEDICAL TESTS

Diane O'Rourke, Timothy Johnson, Seymour Sudman, Richard Warnecke, University of Illinois
Jared B. Jobe, National Center for Health Statistics
Diane O'Rourke, Survey Research Laboratory, University of Illinois,
909 West Oregon, Suite 300, Urbana, IL 61801

KEY WORDS: Regularity, Cancer Screening, Reporting Accuracy

Introduction

This study reports the effect of regularity on the accuracy of reports of three cancer screening tests: Papanicolaou (Pap) smears, mammograms, and breast examinations. Most information about cancer screening in the general public is based on self reports, and there has been concern that the self-reported measures are inaccurate.

Earlier studies using record checks have noted that the overreporting of examinations is in general a serious problem (Gordon, Hiatt, & Lampert, 1993; McKenna et al., 1992; Michielutte et al., 1991; Peters, Bear, & Thomas, 1989; Sawyer et al., 1989; Walter et al., 1988; Warnecke, 1981; Warnecke & Graham, 1976; Warnecke, Havlicek, & Manfredi, 1983). That is, women are aware that they should be getting these tests and, as with voting behavior, may overreport a socially desirable activity. Work by Warnecke and others (Warnecke & Graham, 1976; Warnecke, 1981; Warnecke et al., 1983), however, has also suggested that the accuracy of self-reported Pap smear histories declines among those who do not have regular physical examinations compared with those who regularly receive such exams.

Some research suggests that reporting accuracy may be related to the regularity with which the behavior is experienced. Menon (forthcoming), for example, has demonstrated that the formation and use of schema to make estimates of one's behavior is facilitated by the regularity of that behavior. The use of schema to retrieve information about regular events, as opposed to the counting of individual episodes, is associated with greater accuracy in the recall of the event. Based on this prior research, we hypothesized that more accurate reporting will be found from respondents for whom diagnostic screening tests are conducted regularly than from those tested irregularly, because regular events are easier to schematize accurately.

To test this hypothesis, we analyzed data that were collected by the University of Illinois Survey Research Laboratory in collaboration with the National Center for Health Statistics and RUSH-Anchor HMO in Chicago, for the purpose of evaluating several different questionnaire versions for collecting screening test information (Sudman et al., forthcoming). This paper is limited to testing the hypothesis related to the regularity of diagnostic tests and reporting accuracy.

Methodology

Face-to-Face Interviews

The population examined in this study comprised women aged 50 and older who had been members of the RUSH-Anchor HMO for a minimum of five years. We selected women in this age group because they are the group most likely to misreport having received a Pap smear (Warnecke & Graham, 1976; Warnecke, 1981; Warnecke et al., 1983). Moreover, women over 50 are the primary target population for mammograms. Below this age, this procedure is not recommended for routine casefinding in the absence of symptoms.

After a series of focus groups and think-aloud interviews had been conducted to determine the processes women used to answer questions about health screening tests, face-to-face interviews were conducted with 211 women. These interviews were then compared with the abstracted HMO records for the 178 respondents who gave their permission and for whom validation information was available. The mean age was 59.6 years, 87.6% were African-American, and over three-fourths were employed full- or part-time.

Medical Records Abstraction

Respondent medical records were abstracted, and data for the previous five years were collected for: total number of medical visits, dates that Pap smears and breast examinations were performed, dates mammograms were recommended and performed, and whether or not patients had undergone a full or partial

hysterectomy. Physician progress notes, cytology reports, laboratory reports, and radiology reports were reviewed within the records for information regarding these variables.

We followed the procedure generally adopted in validating self-reports of relying on the medical record as the standard. However, in so doing we recognized that many of those who rely on records have reported that they are an incomplete and not totally reliable source (Demlo, Campbell, & Brown, 1978; Feigl et al., 1988; Romm & Putnam, 1981). A record reabstraction study was thus undertaken to evaluate the reliability of information derived from these medical records (validity is being accessed separately: see Sudman et al., forthcoming). The goal of this reabstraction study was to assess the degree to which complete information was abstracted from medical records. A credentialed medical records technician reabstracted 20 records, and the data were compared with the data initially abstracted. Overall, 98% of the reports showed agreement between the initial and reabstracted data, indicating a very high level of reliability. The matching between the record and the questionnaire was done by a medical records specialist.

Outcome Measures

We used the following response quality measures to summarize respondent reports across study years:

- **Matched data**—the percentage of reports in which the respondent indicated receiving a screening test and the test was verified in the medical records (verified reports/total sample);
- **False reports**—the percentage of reports in which the respondent indicated receiving a screening test for which no matching test was found in the records (unverified reports/total sample);
- **Omissions**—the percentage of reports in which the respondent indicated no test, but a test was found in the records (unreported tests/total sample);
- **No test**—the percentage of reports in which the respondent indicated no test and no evidence of a test was found in the records (verified absences of test/total sample);
- **Gross accuracy**—(Matched data plus no test); this measure is also commonly referred to as an indicator of "concordance;"
- **Percentage reporting**—(Matched data plus false reports);
- **Percentage records**—(Matched data plus omissions); and
- **Ratio**—(Percentage reporting divided by percentage records); we used this as a measure of net bias in test reporting.

It is important to recognize that these measures included multiple reports across years for each individual in the analysis, making the use of traditional tests of significance inappropriate. Consequently, we also developed a total reporting accuracy measure for each respondent for each of the three screening procedures. This measure was a count of the number of concordant reports provided across study years for each screening examination. For each of six years (1987—92), respondents were coded as being accurate if their self-report and medical record were in agreement that the exam in question either was or was not performed during that year (matched data + no test). Respondents were coded as having provided inaccurate reports for each year that they (a) reported receiving a screening procedure that was not confirmed in their medical record (false reports) or (b) reported not receiving a procedure that was identified in their medical record (omissions). Using these data, three outcome measures were produced: the number of accurate, or concordant, reports for each of the three cancer screening procedures that were received over the reporting period. For each procedure, scores ranged from 0 to 6.

We defined regularity as follows: Respondents were classified as regular test takers if there was evidence in the medical record that they were tested every year, every other year, or during four of the six years. Other respondents were classified as irregular, except those whose medical record showed no evidence of cancer screening tests during the study period, who were excluded from the comparison.

Summarized response quality measures are presented in Table 1. Section A reveals that the major difference between regular and irregular test recipients is that the percentage of false reports is significantly higher among irregular test recipients. The cognitive explanation of this finding makes it almost tautological. Many respondents use regularity schema in reporting about these tests. Those respondents whose records indicate that indeed the tests were received regularly would certainly have a lower level of false reports than respondents who did not receive the tests regularly but who thought

they did.

We know of no cognitive reason to have expected that the percentage of omissions would have been significantly different for regular than for irregular test recipients, but it was for breast examinations. One practical reason for this finding could be the lower rate of validity of the medical records for this procedure. Although the record checks for Pap smears and mammograms relied heavily on pathology and radiology reports, respectively, evidence of breast examinations relied solely on the notes of the physicians. It is possible that in some cases, the physician checked off "breast examination" on the record even though one was not done. It is also possible that if one was done, the patient did not recognize or recall it as such. Given the speed with which some physical examinations are conducted, both possibilities are feasible.

Net reporting bias is shown in Section B. Overall, the net biases are far smaller for respondents who received tests regularly for each procedure. They range from +8% for Pap smears to -18% for breast examinations, with mammograms reported with a net bias of only 1% for those getting tests regularly. On the other hand, the net biases for those getting the tests irregularly range from a low of +31% for mammograms to +90% for Pap smears. Detailed data for each test by year show the same results consistently over time; the length of the recall period did not typically affect accuracy.

The proportions of respondents classified as regular test takers for each screening procedure ranged from 31.2% for mammograms to 23.5% and 22.3%, respectively, for breast examinations and Pap smears.

Mean total reporting accuracy by test regularity is shown for each procedure in Table 2. Persons classified as receiving the Pap smear on a regular basis provided a significantly ($p < .001$) greater mean number of accurate reports of this procedure, compared with those classified as receiving this test on only an irregular basis. A similar trend was also observed for mammogram reports, although it did not reach conventional levels of statistical significance ($p = .09$). Breast examination regularity had no effect on the total accuracy with which this procedure was reported.

Because these indicators of total reporting accuracy are highly skewed, each measure was transformed for additional analyses. The mean comparisons reported in Table 2 were reanalyzed using both square root and logit transformations of the total reporting accuracy variables (Mosteller & Tukey, 1977). These transformed variables produced results identical to the initial findings and are therefore not reported here.

Analysis of variance (ANOVA) was next used to assess the effects of test-taking regularity on total reporting accuracy, controlling for potential confounders, including respondent age, education, and perceived health rating (excellent, very good, good, fair, or poor). Additionally, measures of the perceived pain of Pap smears and mammograms were included in the analyses of these procedures, and an indicator of whether or not the respondent had ever experienced a hysterectomy was included in the Pap smear model.

Results for Pap smear total reporting accuracy are presented in Table 3. Test-taking regularity remains a significant predictor of reporting accuracy in this model. No other variables had an independent effect on report accuracy. In Table 4, ANOVA results for breast examination total reporting accuracy are shown. No variables, including test regularity, were associated with accurate reporting of this procedure. As Table 5 indicates, the regularity with which respondents received mammograms, however, approached significance ($p = .09$), along with respondent age ($p = .08$; younger respondents had higher levels of total reporting accuracy). It should be noted that each ANOVA model was reanalyzed using the square root and logit transformations of the total reporting accuracy variables and that these findings supported those presented in Tables 3-5.

Discussion

The accuracy of estimates of frequency of behaviors appears to be influenced by numerous factors (for a review, see Jobe, Tourangeau, & Smith, forthcoming). This paper provides evidence that the regularity with which the behavior is experienced may be one of these factors. As Blair and Burton (1987) have observed, schematic estimation strategies are more likely to be employed when the length of the recall period is longer. Focus groups and think-aloud interviews conducted as precursors to the study reported here confirmed this finding, as virtually all of our respondents were found to use schema to recall cancer screening examinations over a multiyear period (Johnson et al., 1992). When asked

about infrequent events such as these, respondents who experience them on a regular basis (e.g., approximately once per year) can be expected to be more successful in relying on schematic-based recall. Those with irregular event histories will tend to introduce more error into their estimates when using schematic strategies, particularly if no effort is made to adjust their estimates for exceptions.

The use of schema can lead to highly accurate reporting if indeed the events are very regular. Schema, however, will lead to overreporting of behavior if respondents forget to exclude exceptions. (They can also lead to underreporting if respondents forget to include exceptions; see Lessler, Tourangeau, & Salter, 1989.) One might expect that the likelihood of forgetting exceptions would increase with longer time periods, but we saw no evidence of this in this study. If schema are used, then our results would suggest that the order in which questions are asked about details of an event would have no effect on the accuracy of reporting whether the event occurred. The major source of error in reporting of screening tests thus appears to be from those women who report getting a test on a regular basis but who, according to the records, do not.

A major limitation to this study is the use of an HMO sample because women in the sample may be more likely to receive regular screening tests than would women in the general population. It would be useful to have information from a general population sample, although in this case, validation would be more difficult. It is likely, however, that one would continue to find differences in accuracy between regular and irregular test recipients.

References

Blair, E., & Burton, S. (1987). Cognitive processes used by survey respondents in answering behavioral frequency questions. *Journal of Consumer Research*, 14, 280-288.

Demlo, L.K., Campbell, P.M., & Brown, S.S. (1978). Reliability of information abstracted from patients' medical records. *Medical Care*, 16, 995-1005.

Feigl, P., Glaefke, G., Ford, L., Diehr, P., & Chu, J. (1988). Studying patterns of cancer care: How useful is the medical record? *American Journal of Public Health*, 78, 526-533.

Gordon, N.P., Hiatt, R.A., & Lampert, D.I. (1993). Concordance of self-reported data and medical record audit for six cancer screening procedures. *Journal of the National Cancer Institute*, 85, 566-570.

Jobe, J.B., Tourangeau, R., & Smith, A.F. (forthcoming). Contributions of survey research to the understanding of memory. *Applied Cognitive Psychology*.

Johnson, T.P., Jobe, J.B., O'Rourke, D., Sudman, S., & Warnecke, R.B. (1992). Cognitive aspects of reporting cancer prevention examinations and tests. Paper presented at the Annual Meeting of the American Public Health Association, Washington, DC.

Lessler, J.T., Tourangeau, R., & Salter, W. (1989). Questionnaire design in the cognitive research laboratory. *Vital and Health Statistics,* Series 6, No. 1 (DHHS Publication No. PHS 89-1076). Washington, DC: U.S. Government Printing Office.

McKenna, M.T., Speers, M., Mallin, K., & Warnecke, R. (1992). Agreement between patient self-reports and medical records for Pap smear histories. *American Journal of Preventive Medicine*, 8, 287-291.

Menon, G. (forthcoming). Judgments of behavioral frequencies: Memory search and retrieval strategies. In N. Schwarz & S. Sudman (eds.), *Autobiographical Memory and the Validity of Retrospective Reports*. New York: Springer-Verlag.

Michielutte, R., Dignan, M.B., Wells, H.B., Bahnson, J., Smith, M., Wooten, R., & Hale, L.N. (1991). Errors in reporting cervical screening among public health clinic patients. *Journal of Clinical Epidemiology*, 44, 403-408.

Mosteller, F., & Tukey, J.W. (1977). Data Analysis and Regression. Reading, MA: Addison-Wesley.

Peters, R.K., Bear, M.B., & Thomas, D. (1989). Barriers to screening for cancer of the cervix. *Preventive Medicine*, 18, 133-146.

Romm, F., & Putnam, S. (1981). The validity of the medical record. *Medical Care*, 19, 310-315.

Sawyer, J.A., Earp, J.A., Fletcher, R.H., Daye, F.F., & Wynn, T.M. (1989). Accuracy of women's self-report of their last Pap smear. *American Journal of Public Health*, 79, 1036-1037.

Sudman, S., Warnecke, R.B., Johnson, T.P., O'Rourke, D., & Davis, A.M. (forthcoming). Cognitive aspects of reporting cancer prevention examinations and tests. *Vital and Health*

Statistics, Series 6, No. 7 (DHHS Publication No. PHS 94-1083). Washington, DC: U.S. Government Printing Office.

Walter, S.D., Clarke, E.A., Hatcher, J., & Stitt, L.W. (1988). A comparison of physician and patient reports of Pap smear histories. *Journal of Clinical Epidemiology*, 41, 401-410.

Warnecke, R.B. (1981). Intervention in black populations. In C. Mettlin & G.P. Murphy (eds.), *Cancer Among Black Populations* (pp. 167-183). New York: Alan R. Liss.

Warnecke, R.B., & Graham, S. (1976). Characteristics of blacks obtaining Papanicolaou smears. *Cancer*, 37, 2015-2025.

Warnecke, R.B., Havlicek, P.L., & Manfredi, C. (1983). Awareness and use of screening by older-aged persons. In R. Yancik et al. (eds.), *Perspectives on Prevention and Treatment of Cancer in the Elderly* (pp. 275-287). New York: Raven Press.

TABLE 1
RESPONSE QUALITY MEASURES BY TEST REGULARITY
(Percentage)

	Pap smears		Breast examinations		Mammograms	
	Regular	Irregular	Regular	Irregular	Regular	Irregular
Section A						
Gross accuracy	(83.8)	(74.1)	(72.9)	(71.4)	(85.2)	(82.5)
Matched data	67.1	20.1	52.8	24.7	63.9	30.3
No test	16.7	54.0	20.1	46.7	21.3	52.2
False reports	11.0	23.2	6.9	23.2	7.1	14.0
Omissions	5.2	2.7	20.2	5.4	7.7	3.5
Section B						
Percentage reporting	78.1	43.3	59.7	47.9	71.0	44.3
Percentage records	72.3	22.8	73.0	30.1	71.6	33.8
Ratio	1.08	1.90	.82	1.59	.99	1.31
Section C						
N respondents	(35)	(129)	(39)	(131)	(50)	(117)
Reporting years	(6)	(6)	(6)	(6)	(6)	(6)
N reports	(210)	(774)	(234)	(786)	(300)	(302)

TABLE 2
MEAN VALUE OF TOTAL REPORTING ACCURACY BY SCREENING PROCEDURE AND TEST REGULARITY

	Test regularity		
	Regular	Irregular	*t* test
Pap Smears			3.3***
Mean	4.9	4.2	
Standard deviation	1.0	1.4	
N	35	122	
Breast Exams			0.6
Mean	4.4	4.3	
Standard deviation	1.2	1.3	
N	38	124	
Mammograms			1.8+
Mean	5.2	4.9	
Standard deviation		0.9	1.1
N	48	106	

+$p < .10$
***$p < .001$

TABLE 3
ANOVA OF PAP SMEAR TOTAL REPORTING ACCURACY
(N = 153)

	Sum of squares	*df*	Mean square	*F* value
Pap smear regularity	12.9	1	12.9	7.2**
Hysterectomy	0.4	1	0.4	0.2
Age	1.5	1	1.5	0.9
Education	4.1	1	4.1	2.3
Pap smear perceived painful	0.4	1	0.4	0.2
Perceived health rating	1.4	1	1.4	0.8
Residual variance	261.1	140	1.8	
Total variance	281.2	152	1.9	

**$p < .01$

TABLE 4
ANOVA OF BREAST EXAMINATION TOTAL REPORTING ACCURACY
(N = 162)

	Sum of squares	*df*	Mean square	*F* value
Breast exam regularity	0.1	1	0.1	0.1
Age	0.1	1	0.1	0.1
Education	1.2	1	1.2	0.8
Perceived health rating	1.7	1	1.7	0.7
Residual variance	259.1	157	1.7	
Total variance	262.9	161	1.6	

TABLE 5
ANOVA OF MAMMOGRAM TOTAL REPORTING ACCURACY
(N = 152)

	Sum of squares	*df*	Mean square	*F* value
Mammogram regularity	3.1	1	3.1	2.8+
Age	3.4	1	3.4	3.1+
Education	1.0	1	1.0	0.9
Mammogram perceived painful	0.1	1	0.1	0.1
Perceived health rating	0.0	1	0.0	0.0
Residual variance	161.1	146	1.1	
Total variance	168.0	151	1.1	

+ $p < .10$