

A Matter of Opinion about Hysterectomies: Experts' and Practicing Community Gynecologists' Ratings of Appropriateness

ABSTRACT

The degree to which national expert panel survey ratings of the appropriateness of hysterectomy differed from those of a random sample of practicing community gynecologists was determined. Community gynecologists rated hysterectomy as more appropriate on six of eight cervical dysplasia scenarios. Experts agreed among themselves on 19 of 32 indications (intraclass correlation coefficient = 0.66); community gynecologists agreed on 12 of 32 indications (intraclass correlation coefficient = 0.50). Although few differences of opinion existed between experts and community gynecologists, for common clinical scenarios there was a large variation of opinion about the appropriateness of hysterectomy within each group. For areas of clinical uncertainty in which experts' opinions are used in guideline development, additional measures such as process of care, quality of life, and patient preference should be included in discussions about guidelines. (*Am J Public Health*. 1995;85:1125-1128)

Nina A. Bickell, MD, MPH, JoAnne Earp, ScD, Arthur T. Evans, MD, MPH, and Steven J. Bernstein, MD, MPH

Introduction

Practice guidelines are intended to ensure quality and promote the effectiveness of health care.^{1,2} A popular method of guideline development uses group judgments that incorporate expert physician opinions and beliefs.² Reliance on experts' opinions may call into question the judgments of community physicians by implying that a difference exists between the two standards, a discrepancy that could ultimately affect practitioners' acceptance of practice guidelines.³ Although the methodologic validity of expert-derived guidelines has been questioned,⁴ to our knowledge no study has examined the degree to which experts' and community physicians' beliefs about the appropriateness of the same procedures differ. Therefore, we compared two sets of appropriateness ratings for hysterectomy, the second most common surgical procedure in the United States.⁵ One set was given to a national expert managed care panel, and the other to practicing community gynecologists in North Carolina, a state with one of the highest rates (59/100 000) and wide variation (33-103/100 000) in use of hysterectomy.⁶

Methods

Survey Participants

We surveyed two physician groups, one a weighted random sample of 231 North Carolina gynecologists (weighted to include one-third female and two-thirds male gynecologists) who performed at least one hysterectomy in the past year and were licensed by the North Carolina Board of Medical Examiners (n = 671). The second was a group composed of nine physicians nominated by members of the Health Maintenance Organization (HMO) Quality of Care Consortium.⁷

Survey Methods

Forty-one North Carolina gynecologists were excluded because they were not

surgically active or could not be contacted. Of the remaining 190 eligible gynecologists, 140 responded (74% response rate). The nine HMO Quality of Care Consortium panelists completed the North Carolina gynecologists' survey at the conclusion of a 2-day meeting at the RAND Corporation to discuss and rate the appropriateness of their elective indications for hysterectomy.^{8,9}

Survey Questionnaire

The questionnaire asked about personal and practice characteristics and the appropriateness of performing hysterectomy for 32 clinical scenarios. Sixteen scenarios covered the most prevalent indication for hysterectomy (fibroids); eight concerned a common equivocal indication (cervical dysplasia); four covered a common subjective indication in North Carolina (cancer fear); and four were on cancer.⁶ All scenarios concerned women who did not desire future fertility and who differed by age (35 vs 45 years old) and by disease severity.

Appropriateness was considered in tertiles, with median scores of 1 through 3 rated as inappropriate (risks outweigh benefits), 3.5 through 6 as equivocal (risks and benefits roughly equal), and 6.5 through 9 as appropriate (benefits outweigh risks).

At the time of this study, Nina Bickell was with the Department of Medicine, University of North Carolina, Chapel Hill, NC. JoAnne Earp is with the Department of Health Behavior and Health Education, School of Public Health, University of North Carolina, Chapel Hill. Arthur T. Evans is with the Department of Medicine, University of North Carolina, Chapel Hill. At the time of this study, Steven J. Bernstein was with the Department of Medicine, University of California, Los Angeles, and the RAND Corporation, Santa Monica, Calif.

Requests for reprints should be sent to Nina A. Bickell, MD, MPH, Department of Health Policy, Mt. Sinai Medical Center, Box 1077, One Gustave L. Levy Place, New York, NY 10029-6574.

This paper was accepted January 11, 1994.

TABLE 1—Hysterectomy Appropriateness: Median Scores and Agreement within Physician Groups

Clinical Scenario	Expert Median (SD)		NC Gynecologist Median (SD)	
	35 y	45 y	35 y	45 y
Stage 1 endometrial cancer	9 ^A (0)	9 ^A (0)	9 ^A (0.9)	9 ^A (0.9)
Stage 1 cervical cancer	9 ^A (0)	9 ^A (0)	9 ^{NA} (2.4)	9 ^{NA} (2.4)
Dysplasia on Pap smear*	1 ^A (0.7)	1 ^A (0.3)	1 ^{NA} (2.5)	1 ^{NA} (2.7)
Dysplasia on punch biopsy*	1 ^A (2.7)	1 ^A (2.6)	5 ^{NA} (2.8)	5 ^{NA} (3.0)
Dysplasia on cone biopsy*	2.5 ^{NA} (3.2)	3 ^{NA} (3.2)	8 ^{NA} (2.4)	8 ^{NA} (2.2)
Cervical carcinoma in situ*	4 ^{NA} (2.5)	4 ^{NA} (2.5)	8 ^{NA} (2.0)	8 ^{NA} (1.9)
Cancer phobia without risk factors**	1 ^A (1.3)	1 ^A (1.7)	1 ^{NA} (1.8)	2 ^{NA} (2.0)
Cancer phobia with risk factors	2 ^{NA} (2.5)	2 ^{NA} (2.5)	3.5 ^{NA} (2.4)	4 ^{NA} (2.5)
Bleeding fibroids with anemia refractory to iron therapy*	8 ^A (1.2)	8 ^A (1.1)	9 ^A (1.2)	9 ^A (1.2)
Bleeding fibroids in a nonanemic woman	5 ^{NA} (0.9)	6 ^{NA} (0.9)	6 ^{NA} (1.1)	6 ^{NA} (1.0)
Bleeding fibroids refractory to conservative surgical procedures	9 ^A (2.1)	9 ^A (2.4)	8 ^A (2.1)	9 ^A (2.0)
Bleeding fibroids with anemia responding to iron	5 ^{NA} (2.2)	6 ^{NA} (2.3)	7 ^{NA} (2.3)	7 ^{NA} (2.3)
Painful fibroids in an otherwise healthy woman	7 ^{NA} (2.2)	6 ^{NA} (2.1)	8 ^{NA} (1.6)	7 ^{NA} (1.7)
Painful fibroids refractory to medical therapy and psychological counseling	9 ^A (1.6)	8 ^A (1.3)	9 ^A (1.1)	9 ^A (1.2)
Painful fibroids refractory to conservative surgical procedures	9 ^A (1.1)	8 ^A (0.9)	9 ^A (1.1)	9 ^A (1.1)
Painful fibroids refractory to oral medical therapy**	9 ^{NA} (1.2)	7 ^A (0.9)	9 ^A (1.1)	9 ^A (1.1)

Note. Agreement was defined as all scores falling within tertiles of 1 through 3 (inappropriate), 3.5 through 6 (equivocal), or 6.5 through 9 (appropriate) after the top and bottom 10% of sample were discarded for the North Carolina physicians and the high and low scores were discarded for the HMO expert panel. ^A = Agreement; ^{NA} = no agreement.

* $P < .05$ by Wilcoxon rank sum test between experts and NC gynecologists for scenarios with 35- and 45-yr-old women.

** $P < .05$ by Wilcoxon rank sum test between experts and NC gynecologists for scenarios with 45-year-old women.

Statistical Analysis

We present median values because the expert group was small and mean values might be misleading. Overall agreement between the experts and the community gynecologists was assessed with the Wilcoxon rank sum statistic. Agreement within each group was assessed by an intraclass correlation coefficient to determine whether each group agreed more with its own members than could be accounted for by chance alone and by a second method that allowed, after discarding one extremely high rating and one extremely low rating, an examination of whether the remaining seven ratings all fell within the 3-point tertile containing the median. This RAND method of determining agreement was also used with the community physicians by discarding the top and bottom 10% of ratings and then assessing whether the remaining

80% fell within one tertile of appropriateness.

Results

The median number of hysterectomies performed per month was three for both experts and community gynecologists. Twenty-two percent of the experts and 8% of the community gynecologists had completed subspecialty training. A much higher percentage of community gynecologists' patients had private (fee for service) insurance compared with the managed care experts' patients (70% vs 10%; $P < .0001$). On average, experts and community gynecologists graduated from residency training in 1976.

Appropriateness

Overall, little difference existed between experts' and community gynecolo-

gists' appropriateness ratings. Both groups rated hysterectomy as appropriate for all 4 cancer (cervical and endometrial) scenarios and for 13 of 16 fibroid scenarios (Table 1). There was a statistically significant difference between experts and community gynecologists regarding the appropriateness of hysterectomy for all cervical dysplasia scenarios; for six of eight such scenarios community gynecologists rated hysterectomy more appropriate for women with carcinoma in situ or significant dysplasia on punch or cone biopsy (Figure 1).

Agreement

The finding that few differences exist between the experts and community gynecologists, however, depends on the variation in appropriateness scores within each group. Using the RAND tertile method for assessing agreement, we found that the experts agreed among themselves that hysterectomy was appropriate for all endometrial and cervical cancer scenarios (4/4) and for 56% (9/16) of the fibroid indications (Table 1). They agreed about the inappropriateness of hysterectomy for women with cancer fear but without cancer risk factors for two of four scenarios and for women whose Pap smear or cervical punch biopsy showed dysplasia for four of eight scenarios. The experts did not agree for nonanemic women with bleeding fibroids, women with bleeding fibroids and iron-responsive anemia, women who feared cancer and had cancer risk factors, women with dysplasia on cone biopsy, and those with cervical carcinoma in situ. Overall, experts agreed among themselves on 19 of 32 indications (59%), with the intraclass correlation coefficient of 0.66 showing a relatively good amount of agreement.

We found that North Carolina community gynecologists agreed among themselves about the appropriateness of hysterectomy for endometrial cancer and for 10 of 16 fibroid scenarios (63%). They did not agree with each other for either of the two cervical cancer scenarios, the four fear-of-cancer scenarios, eight cervical dysplasia scenarios, and four bleeding-fibroid scenarios in nonanemic and iron-responsive anemic women. Overall, North Carolina community gynecologists agreed on only 12 of 32 clinical scenarios (38%), with an intraclass correlation coefficient of 0.50 showing only a moderate level of agreement beyond chance alone.

Discussion

Even after 2 days of discussion about the appropriateness of hysterectomy for

various elective indications, experts did not agree among themselves regarding many common scenarios. They agreed, however, more than community gynecologists, who rated the same scenarios without a discussion of their differences of opinion about the procedure. Although community gynecologists rated hysterectomy as more appropriate for cervical dysplasia than did the experts, few other differences existed between these groups about the appropriateness of hysterectomy. Previous reports found that multiple rounds of ratings and/or discussion lead to greater agreement.⁹ Discussion among groups of physicians and evaluation of patient outcomes for clinical areas with large practice variation and known differences of opinion may be helpful in reducing disagreement.

Those scenarios about which expert panel members agreed on appropriateness but community gynecologists did not may represent areas in which expert-derived guidelines could have the greatest impact on community physician practice. Although explicit criteria may help minimize practice variation for agreed-on scenarios, such criteria have less impact on varying opinions for clinically uncertain scenarios, a limitation of opinion-derived guidelines. In our list of indications, these uncertain scenarios included women with bleeding fibroids responsive to treatment and women with moderate to severe cervical dysplasia, two common conditions in North Carolina for which up to 45% of all hysterectomies are performed.⁵

The small level of disagreement between experts' and community physicians' appropriateness ratings belies the large variation in opinion that exists within each group. This variation may reflect uncertainty about a procedure often done electively to improve quality of life.¹⁰ Treatment choice and appropriateness ratings are affected by clinical and nonclinical factors including social characteristics, patient preference, quality of life, and process of medical care.¹¹⁻¹⁵ For areas of clinical uncertainty, these non-clinical factors likely exert a greater influence on decision making. Ideally, empirical research should be used to guide decision making. A recent outcomes study of symptom relief and patient satisfaction among women with bleeding fibroids treated medically or surgically provides an example of how information can be used for treatment discussions between patients with symptomatic fibroids and their gynecologists.^{16,17} How-

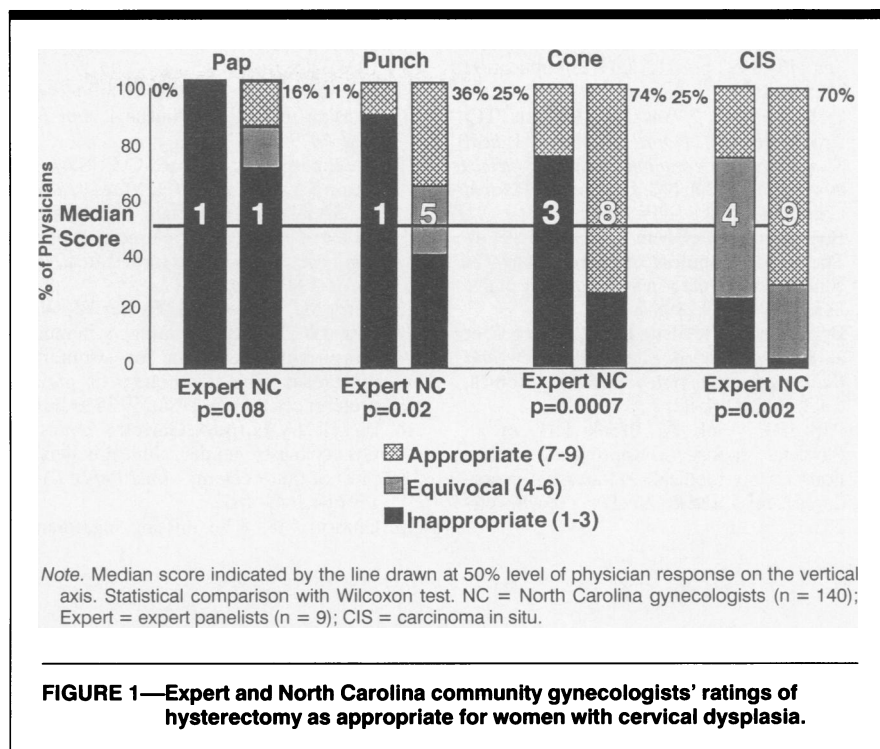


FIGURE 1—Expert and North Carolina community gynecologists' ratings of hysterectomy as appropriate for women with cervical dysplasia.

ever, in the absence of such evidence, opinion-derived guidelines provide some framework for decision making. Although this approach may not be optimal, it gives a baseline reference point, a description of current practice standards, and points out areas for future research needed to provide the evidence and rationale for clinical areas of uncertainty.

Expert panelists were less likely to rate hysterectomies as appropriate than were North Carolina community gynecologists. This may have occurred because the experts were all HMO clinicians, and a goal of HMOs is to minimize inappropriate procedures. In addition, practice styles vary by region,¹⁸ and the experts were from all four major US census regions. Because the South has the highest hysterectomy rate, regionality may have contributed to the North Carolina community gynecologists' higher ratings of hysterectomy appropriateness.

Given our strict definition of agreement, we would expect more disagreement among a larger group because there is a greater likelihood that a single individual's ratings will fall outside the median tertile. Although we did find more disagreement in the larger group, in only 2 of the 32 scenarios was the disagreement due to a few community gynecologists' ratings falling outside the median tertile.

Expert opinion-derived guidelines are one method for providing a coherent approach to decisions about medical care;

we have shown that problems with such guidelines may arise in areas of marked clinical uncertainty. Scenarios about which neither experts nor community physicians can agree should provide the focus for future clinical studies.¹⁹ Process of care, quality of life, and patient preference measures should be included in discussions about appropriateness and guideline development,¹³ particularly for clinically uncertain conditions. □

Acknowledgments

Dr Bickell was a Robert Wood Johnson Clinical Scholar at the time of this study.

The authors gratefully acknowledge the contributions of Drs Robert Brook and Gordon DeFries in making this project possible, and Dr Albert Siu's thoughtful comments.

References

1. Brook RH. Practice guidelines and practicing medicine. Are they compatible? *JAMA*. 1989;262:3027-3030.
2. Audet AM, Greenfield S, Field M. Medical practice guidelines: current activities and future directions. *Ann Intern Med*. 1990;113:709-714.
3. Lomas J, Anderson GM, Dominick-Pierre K, Vayda E, Enkin MW, Hannah WJ. Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N Engl J Med*. 1989;321:1306-1311.
4. Phelps CE. The methodologic foundations of studies of the appropriateness of medical care. *N Engl J Med*. 1993;329:1241-1245.
5. Graves EJ. National Center for Health

Statistics, national hospital discharge survey: annual summary, 1990. *Vital Health Stat [13]*. 1992 no. 112. DHHS publication PHS 92-1773.

6. DeFries GH, Evans AT, Ricketts TC, Cromartie EP. *North Carolina Medical Society Practice Variation Study of Hysterectomy*. Chapel Hill, NC: University of North Carolina; 1989.
7. Bernstein SJ, McGlynn EA, Siu AL, et al. The appropriateness of hysterectomy. A comparison of care in seven health plans. *JAMA*. 1993;269:2398-2402.
8. Bernstein SJ, McGlynn EA, Kamberg C, et al. *Hysterectomy: A Literature Review and Ratings of Appropriateness*. Santa Monica, Calif: RAND; 1992.
9. Park RE, Fink A, Brook RH, et al. Physician ratings of appropriate indications for six medical and surgical procedures. *Am J Public Health*. 1986;76:766-772.
10. Carlson KJ, Nichols DH, Schiff I. Indications for hysterectomy. *N Engl J Med*. 1993;328:856-860.
11. Eisenberg J. The sociological influences on decision making by clinicians. *Ann Intern Med*. 1979;90:957-964.
12. Greenberg ER, Chute CG, Stukel T, Baron JA, Freeman DH, Yates J, Korson R. Social and economic factors in the choice of lung cancer treatment. A population-based study in two rural states. *N Engl J Med*. 1988;318:612-617.
13. Barry MJ, Mulley AG, Fowler FJ, Wennberg JW. Watchful waiting vs immediate transurethral resection for symptomatic prostatism. The importance of patients' preferences. *JAMA*. 1988;259:3010-3017.
14. Bickell NA, Earp JA, Garrett J, Evans AT. Gynecologists' gender, clinical beliefs and rates of hysterectomy. *Am J Public Health*. 1994;84:1649-1652.
15. Chassin MR. The missing ingredient in health reform: quality of care. *JAMA*. 1993;260:377-378.
16. Carlson KJ, Miller BA, Fowler FJ. The Maine Women's Health Study, I: outcomes of hysterectomy. *Obstet Gynecol*. 1994;83:556-565.
17. Carlson KJ, Miller BA, Fowler FJ. The Maine Women's Health Study, II: outcomes of nonsurgical management of leiomyomas, abnormal bleeding, and chronic pelvic pain. *Obstet Gynecol*. 1994;83:566-572.
18. Leape LL, Park RE, Kahan JP, Brook RH. Group judgments of appropriateness: the effect of panel composition. *Qual Assur Health Care*. 1992;44:151-159.
19. Naylor CD, Sibbald WJ, Sprung CL, Pinfold SP, Calvin JE, Cerra FB. Pulmonary artery catheterization. Can there be an integrated strategy for guideline development and research promotion? *JAMA*. 1993;269:2407-2411.

Hormonal Replacement Therapy and Morbidity and Mortality in a Prospective Study of Postmenopausal Women

Aaron R. Folsom, MD, Pamela J. Mink, Thomas A. Sellers, PhD, Ching-Ping Hong, MS, Wei Zheng, MD, PhD, and John D. Potter, PhD, MBBS

ABSTRACT

We assessed the association of hormonal replacement therapy with mortality and incidence of multiple diseases in over 40 000 postmenopausal women followed for 6 years as part of the Iowa Women's Health Study. Compared with women who never used hormone replacement therapy, current users had multivariate adjusted relative risks (RR) as follows: total mortality (RR = 0.78; 95% confidence interval [CI] = 0.65, 0.94), coronary heart disease mortality (RR = 0.74; 95% CI = 0.48, 1.12), endometrial cancer incidence (RR = 4.3; 95% CI = 2.7, 6.9), breast cancer incidence (RR = 1.23; 95% CI = 0.99, 1.55), colon cancer incidence (RR = 0.72; 95% CI = 0.46, 1.12), and hip fracture incidence (RR = 0.53; 95% CI = 0.31, 0.91). (*Am J Public Health*. 1995;85:1128-1132)

Introduction

In the absence of large-scale randomized clinical trial data, the risks and benefits of hormone replacement therapy for postmenopausal women remain unresolved.^{1,2} A recent meta-analysis of observational studies estimated that hormone replacement therapy use carries reduced relative risks of 0.65 for coronary heart disease and 0.75 for fracture, but relative risks of 0.96 for stroke and 1.25 for breast cancer.² Relative risk estimates for endometrial cancer were 8.22 for estrogen therapy and 1.00 for combination estrogen plus progestin therapy,² although few data on combination hormone replacement therapy are available. Although the clinical trial component of the Women's Health Initiative³ should help resolve the risk-benefit controversy, results will not be available for over a decade and it cannot address the long-term risks of hormone replacement therapy. In the meantime, additional observations on hormone replacement therapy from contemporary cohort studies with multiple disease end points may be of use to health professionals.

Methods

The Iowa Women's Health Study cohort^{4,5} consists of 41 837 women aged 55 through 69 years who had a valid Iowa driver's license in 1985. A mailed questionnaire provided information on current and past hormone replacement therapy, prevalent diseases, anthropometric data, and other risk factors. Hormone replacement therapy was determined by the following questions: "Have you ever used pills other than birth control pills which contained estrogen or other female hormones (e.g., at the change of life or menopause, after surgery, or at another time)?" and, if yes, "How long did you take estrogens or other female hormone pills (other than birth control pills)?"

The authors are with the Division of Epidemiology, School of Public Health, University of Minnesota, Minneapolis.

Requests for reprints should be sent to Aaron R. Folsom, MD, Division of Epidemiology, School of Public Health, University of Minnesota, 1300 S Second St, Suite 300, Minneapolis, MN 55454-1015.

This paper was accepted December 15, 1994.