A RAND NOTE

Does Inappropriate Use Explain Geographic Variations in the Use of Health Care Services?  
A Study of Three Procedures

Mark R. Chassin, Jacqueline Kosecoff, R.E. Park,  
Constance M. Winslow, Katherine L. Kahn,  
Nancy J. Merrick, Joan Keesey, Arlene Fink,  
David H. Solomon, Robert H. Brook

November 1987
The research described in this report was sponsored by the Commonwealth Fund, the John A. Hartford Foundation, the U.S. Health Care Financing Administration, U.S. Department of Health and Human Services, the Pew Memorial Trust, and the Robert Wood Johnson Foundation.

This Note contains an offprint of RAND research originally published in a journal or book. The text is reproduced here, with permission of the original publisher.

The RAND Publication Series: The Report is the principal publication documenting and transmitting RAND's major research findings and final research results. The RAND Note reports other outputs of sponsored research for general distribution. Publications of The RAND Corporation do not necessarily reflect the opinions or policies of the sponsors of RAND research.
Does Inappropriate Use Explain Geographic Variations in the Use of Health Care Services?  
A Study of Three Procedures

Mark R. Chassin, Jacqueline Kosecoff, R.E. Park,  
Constance M. Winslow, Katherine L. Kahn,  
Nancy J. Merrick, Joan Keesey, Arlene Fink,  
David H. Solomon, Robert H. Brook

November 1987

Prepared for  
The Commonwealth Fund  
The John A. Hartford Foundation  
The U.S. Health Care Financing Administration,  
U.S. Department of Health and Human Services  
The Pew Memorial Trust  
The Robert Wood Johnson Foundation
Does Inappropriate Use Explain Geographic Variations in the Use of Health Care Services?

A Study of Three Procedures

Mark R. Chassin, MD, MPP, MPH; Jacqueline Kosecoff, PhD; R.E. Park, PhD; Constance M. Winslow, MD; Katherine L. Kahn, MD; Nancy J. Merrick, MD, MSPH; Joan Keesey; Arlene Fink, PhD; David H. Solomon, MD; Robert H. Brook, MD, ScD

We studied the appropriateness of use of coronary angiography, carotid endarterectomy, and upper gastrointestinal tract endoscopy and its relationship to geographic variations in the rates of use of these procedures. We selected geographic areas of high, average, and low use of these procedures and randomly sampled Medicare beneficiaries who had received one of the procedures in 1981. We determined the indications for the procedures using a detailed review of medical records and used previously developed ratings of appropriateness to assign an appropriateness score to each case. Differences among sites in levels of appropriateness were small. For example, in the high-use site for coronary angiography, 72% of the procedures were appropriate, compared with 81% in the low-use site. Coronary angiography was performed 2.3 times as frequently in the high-use site compared with the low-use site. Under the conditions of this study, we did find significant levels of inappropriate use: 17% of cases for coronary angiography, 32% for carotid endarterectomy, and 17% for upper gastrointestinal tract endoscopy. We conclude that differences in appropriateness cannot explain geographic variations in the use of these procedures.

Many explanations for these differences have been proposed. These range from differences among geographic areas in the number of available physicians or hospital beds to differences in the sociodemographic composition of patient populations to differences in the degree of uncertainty facing physicians regarding the use of different procedures.

We were concerned about whether such geographic variations might be explained by differences in the appropriateness with which physicians use medical and surgical procedures. The hypothesis was that a high rate of inappropriate use might accompany and explain a high overall use rate in a particular geographic area. We report herein the results of a study in which we measured how appropriately physicians performed coronary angiography, carotid endarterectomy, and upper gastrointestinal (GI) tract endoscopy in the elderly Medicare population in 1981 in areas of the United States characterized by high, average, and low use of the procedures.

METHODS
Site Selection

We have previously described the methods by which we measured population-based rates of use of health care services in our 15 sites. Briefly, we obtained complete files of physician claims for the year 1981 from Medicare insurance carriers in Arkansas, Colorado, Iowa, Massachusetts, Montana, Pennsylvania, South Carolina, and northern California. We divided both northern California and Pennsylvania into three smaller areas and Massachussetts into two smaller areas, choosing the area boundaries to minimize the number of patients that crossed them to receive health care services.

We calculated population-based rates of use for 153 procedures (accounting for 87% of total physician charges to Medicare in these sites) by counting the frequency with which physicians practicing in each site provided services to Medicare beneficiaries who resided therein. We calculated rates by dividing this count by the number of Medicare part B enrollees who lived in each site in 1981. We excluded claims for patients living outside an area who came into it for services, as well as services performed for residents of an area by physicians practicing outside the area. Rates were adjusted for age and sex differences among sites. Statistically significant differences were demonstrated among sites for all procedures.

Table 1 gives the rates of use in the sites we selected for the study reported herein. For each procedure, we selected one high-use site and one low-use site. The selection of the third study site was
based on a different strategy for each procedure.

For coronary angiography we studied two low-use sites, because they represented geographic areas with notably different characteristics (site 3 is a relatively rural site, while site 2 is considerably more urban). For carotid endarterectomy, we chose a site with an average rate so that at least one of our two sites would not represent an extreme rate. For upper GI tract endoscopy, we chose two high-use sites, because different kinds of physicians reported doing the endoscopies in the two sites; in site 2, physicians who described themselves to the insurance carrier as gastroenterologists performed 72% of the endoscopies, whereas in site 5, internists performed 61%.

**Procedure Selection and Appropriateness Rating**

From among all the health care services for which we found significant variation in use rates among sites, we selected the three listed above for this study. These procedures were selected because each is frequently performed in the Medicare population and each consumes substantial resources.

For each of these procedures we prepared reviews of the medical literature that summarized existing knowledge concerning the circumstances under which each procedure has been proved to be efficacious, its risks, its costs, and opinions concerning when it ought to be used. From these reviews we developed comprehensive and detailed lists of indications or specific clinical situations in which the use of each procedure might be considered.

Panels of expert physicians rated each indication on a nine-point scale of appropriateness (where 9 indicates extremely appropriate; 5, equivocal; and 1, extremely inappropriate). We defined appropriateness as present to the extent that the expected health benefits of a procedure exceed its expected negative consequences by a sufficiently wide margin that the procedure is worth doing. We explicitly excluded considerations of monetary costs from this definition. Panelsists were given our literature reviews and specifically instructed to rate indications for the performance of these procedures in 1981.

The process by which the panels were constituted and conducted has been reported elsewhere. Briefly, each panel rated and revised the initial set of indications by mail, then met at The Rand Corporation, Santa Monica, Calif, reviewed and finalized the list of indications, reviewed summaries of first-round ratings, discussed the clinical areas in which disagreement among panelists was prominent, and "re-rated" all indications. The lists of final, rated indications are available elsewhere.

(Reports are available at nominal cost on request from The Rand Corporation.) Final indications rated by the panels number 300 for coronary angiography, 984 for carotid endarterectomy, and 1058 for upper GI tract endoscopy.

Our consensus method did not force panelists to reach agreement on appropriateness. As we have described elsewhere, significant degrees of disagreement for specific indications persisted after the meetings. We considered indications appropriate if their median ratings were from 7 to 9 without disagreement. We considered indications inappropriate if their median ratings were from 1 to 3 without disagreement. We considered indications equivocal if their median ratings were from 4 to 6, or if panelists disagreed on the appropriateness of performing the procedure for that specific indication. For this purpose, we defined disagreement as occurring when at least three panelists rated the indication appropriate (ratings of 7 to 9) and at least three panelists rated the indication inappropriate (ratings of 1 to 3), regardless of the median rating.

**Data Collection**

We selected a random sample of cases of each of the three study procedures in each study site. These samples were drawn directly from the claim data that had been used to calculate each site's rates of utilization. We have described in detail in the next article in this series the complex, three-tiered approach we employed to recruit the participation of this community-based sample of 913 physicians who performed the 5411 study procedures. Ninety percent, or 819 physicians, participated in the study; they performed 4988 (92%) of all sampled procedures. We recruited the participation of 227 (99%) of the 230 hospitals at which the sampled procedures had been performed. We have also described our data collection methods, as well as the methods and results of the studies we did of the validity of the indications assigned to each case by these methods. We obtained complete clinical data on 96% of eligible cases.

**Statistical Methods**

We sampled roughly equal numbers of cases for each procedure in each site, to maximize the power to detect differences among sites. With approximately
Table 3.— Appropriateness of Use of Three Procedures by Site With 95% Confidence Intervals (CI)

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Site No.</th>
<th>Sample Size, No.</th>
<th>Appropriate (95% CI)</th>
<th>Equivocal (95% CI)</th>
<th>Inappropriate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronary angiography</td>
<td>1</td>
<td>628</td>
<td>71.7 (68.1-75.2)</td>
<td>10.0</td>
<td>18.3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>540</td>
<td>60.2 (57.2-53.4)</td>
<td>6.6</td>
<td>13.6</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>535</td>
<td>80.7 (77.4-84.1)</td>
<td>4.5</td>
<td>14.8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>1677</strong></td>
<td><strong>74.0</strong></td>
<td><strong>8.5</strong></td>
<td><strong>17.4</strong></td>
</tr>
<tr>
<td>Carotid endarterectomy</td>
<td>1</td>
<td>600</td>
<td>36.5 (32.7-40.4)</td>
<td>33.9</td>
<td>29.6</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>492</td>
<td>29.7 (25.6-33.7)</td>
<td>30.1</td>
<td>40.2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>210</td>
<td>41.9 (35.2-48.6)</td>
<td>59.0</td>
<td>29.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>1302</strong></td>
<td><strong>35.3</strong></td>
<td><strong>32.3</strong></td>
<td><strong>32.4</strong></td>
</tr>
<tr>
<td>Upper GI tract endoscopy*</td>
<td>5</td>
<td>509</td>
<td>70.9 (67.0-74.9)</td>
<td>11.2</td>
<td>17.9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>462</td>
<td>72.9 (68.9-77.0)</td>
<td>7.6</td>
<td>19.5</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>614</td>
<td>71.6 (68.1-75.2)</td>
<td>13.9</td>
<td>14.9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>1585</strong></td>
<td><strong>72.0</strong></td>
<td><strong>10.8</strong></td>
<td><strong>17.2</strong></td>
</tr>
</tbody>
</table>

*GI indicates gastrointestinal.

Table 4.— Most Frequent Appropriateness Indications for Three Procedures

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Indication</th>
<th>% of All Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronary angiography</td>
<td>Patient's primary cardiac abnormality is valvular disease. Patient hospitalized for unstable angina. Pain is controlled during admission, angiography performed during admission</td>
<td>18</td>
</tr>
<tr>
<td>Coronary angiography</td>
<td>Patient has had multiple carotid transient ischemic attacks, has never received medical therapy, has medium surgical risk, and 70%-99% stenosis of artery ipsilateral to symptomatic hemisphere. Patient is asymptomatic, has low stroke risk, low surgical risk, and 70%-99% stenosis of the artery operated on.</td>
<td>6</td>
</tr>
<tr>
<td>Coronary angiography</td>
<td>Patient has dysphagia, previous barium swallow shows an anatomic lesion. Patient has had previous roentgenographic examination of the upper GI tract within 3 mo that suggests a malignant lesion.</td>
<td>12</td>
</tr>
</tbody>
</table>

*Our panel defined a carotid transient ischemic attack as any transient unilateral neurologic deficit lasting less than 24 hours. We defined medical therapy as any antiocoagulant or platelet-inhibiting agent at any dose. We defined surgical risk as low, medium, or high according to the criteria defining the Goldman index, the Dripps-American Society of Anesthesiologists criteria, or both. Stroke risk was defined as normal if the risk of stroke was less than 10% within eight years, using data from the Framingham study. A full and detailed list of definitions and indications is available elsewhere.16

500 sample cases for each procedure in each site, we could expect to detect a true difference of eight percentage points in appropriate or inappropriate use in at least 80% of repeated trials.

Although the sample sizes were approximately equal, the population sizes differed greatly. For some procedures in some sites, all of the population cases were sampled; for one procedure in one site, fewer than 10% of the population cases were sampled. To obtain unbiased estimates of appropriate or inappropriate use across all sites, we used inverse sample-probability weights when we calculated statistics for all sites.

We tested for differences in the distribution of appropriateness ratings across sites using χ² tests on unweighted sample frequencies.26

RESULTS

Table 2 gives the appropriateness of use of each of our three study procedures in each of the geographic sites in which we collected clinical data. For each procedure, it lists the percent of cases in each site in each of the three appropriateness categories: appropriate, equivocal, and inappropriate. It also gives the rate of use of each procedure in each site as determined by our previous analysis of Medicare claims data. More detailed data on appropriateness are presented in Table 3, including confidence intervals on all estimates of appropriateness and one additional significant figure on all estimates.

We found small, but statistically significant differences in appropriateness among the three sites studied for each procedure. For coronary angiography, the highest-use site had the lowest rate of appropriateness (72%) and the lowest-use site had the highest rate of appropriateness (81%). For carotid endarterectomy, there were no significant differences between the high- and low-use sites. The average-use site, however, had a significantly higher rate of inappropriate procedures (40%) than either of the other two (30% and 29%, respectively). For upper GI tract endoscopy, the low-use site had the lowest rate of inappropriate procedures (15%), compared with the two sites with higher rates of use (18% and 19%, respectively).

As given in Table 2, across all sites we found that coronary angiography was performed appropriately 74% of the time, carotid endarterectomy, 55%, and upper GI tract endoscopy, 72%. Overall, the three procedures were performed for equivocal indications in 9%, 32%, and 11% of cases, respectively. Finally, inappropriate cases accounted for 17% of coronary angiographies across all sites, 32% of carotid endarterectomies, and 17% of upper GI tract endoscopies.

Table 4 lists the two most frequent appropriate indications for each procedure across all three sites and the percent of cases of each procedure accounted for by each. Table 5 does the same for the most frequent equivocal indications. Table 6 lists the most frequent inappropriate indications.

COMMENT

We measured the appropriateness of performing three procedures in a randomly selected, community-based sample of cases from five sites across the United States in 1981. The sites were selected because they represented high, average, or low use for each of the procedures studied. We found small, but statistically significant differences in appropriateness among sites in each of the procedures we studied.

Although these differences were in the direction supporting the hypothesis (ie, less appropriate or more inappropriate procedures in areas of high use), in no case can differences in appropriateness explain the large differences in overall rates previously reported. Thus, we did not find evidence to support the hypothesis that areas with high use of medical and surgical procedures show these high rates primarily or to any
meaningful extent because physicians in these areas perform procedures more often for inappropriate indications than their counterparts in areas of lower use.

While further research is needed to explain geographic variations in health care service use, we believe our data suggest several new approaches. First, we must begin to examine the question of underuse more scientifically. Our study design did not permit us to measure directly the extent of underuse, if any, and its contribution to geographic differences. Second, we should also begin to look more closely at regional differences in disease incidence as possible partial explanatory factors. Third, regional social and cultural differences may produce differences among geographic areas in the stage of illness at which patients seek health care. And fourth, we should give more attention to the role of primary care physicians in determining which patients eventually receive specialized procedures such as the ones we studied. Openness of referral channels and the threshold for referral might be important factors in explaining regional differences in utilization rates for procedures.

We believe our method of cataloging and rating indications for these procedures using expert physician panels has produced valid data on appropriateness. We have discussed this issue in detail elsewhere for carotid endarterectomy.14 We also believe the data in Tables 4 to 6 provide further evidence that this method does distinguish clinically meaningful differences in appropriateness. For example, the data in Table 4 show that the most frequent appropriate indication for upper GI tract endoscopy was for patients with dysphagia in whom a previous upper GI tract roentgenographic study had shown anatomic lesion, such as an esophageal stricture or ulcer. Few, if any, physicians would deny that this is a clearly appropriate indication for upper GI tract endoscopy.

The indication listed in Table 5 as the most frequently occurring equivocal indication for upper GI tract endoscopy would be more likely to provoke controversy. Under what circumstances should patients with uncomplicated peptic symptoms who have received inadequate medical therapy and who have not had upper GI tract roentgenographic studies receive endoscopy? It seems likely that physicians will differ (as did our panel) far more on this issue than on the appropriateness of the procedure for patients with dysphagia.

Slightly, the most frequent inappropriate indication for upper GI tract endoscopy, listed in Table 6, is clearly less appropriate than the equivocal indication. Should patients with uncomplicated peptic symptoms and normal upper GI tract roentgenographic studies who have received inadequate medical therapy ever undergo upper GI tract endoscopy before receiving a trial of adequate medical treatment? Perhaps there are rare special circumstances that might justify the procedure under these conditions, but we believe most physicians would agree that ordinarily endoscopy is inappropriate.

In two of the three procedures, we found most use to be for clearly inappropriate indications over all sites: 74% for coronary angiography and 72% for upper GI tract endoscopy. For carotid endarterectomy, however, 35% of procedures across all sites were performed for clearly appropriate indications. Carotid endarterectomy was again different, with 32% of cases performed for equivocal indications.

In all three procedures we found substantial numbers of procedures performed for inappropriate indications. Over all sites, 17% of coronary angiographies, 17% of upper GI tract endoscopies, and 32% of carotid endarterectomies were inappropriate, judged against our expert panels' ratings.

These findings are especially worrisome, because we constructed our method of assigning indications to give cases higher appropriateness ratings where information was ambiguous. Many of the definitions we used leaned in the same direction. For example, our definition of maximal medical therapy for angina was met if the patient was receiving any long-acting nitrate and any β-blocker in any dosage schedule. Many clinicians might regard this definition as insufficiently stringent. Similarly, we classified any transient unilateral neurologic deficit as a carotid transient
ischemic attack no matter how vaguely described. If differing opinions about degree of carotid stenosis were recorded in the medical record, we used the highest recorded degree of stenosis. Our panel's definition of maximal medical therapy for peptic symptoms was met if the patient was receiving any antacid or any H₂-receptor antagonist, regardless of daily dosage schedule. All of these decisions tended to result in higher ratings of appropriateness than would the application of stricter standards. The results of our validity studies are also consistent with this observation.³

Our definition of appropriateness excluded consideration of cost from the negative consequences of performing procedures because we wanted the definition to be a purely medical one. Had we asked our panelists to include cost considerations, their ratings of appropriateness might have been lower.

Several factors might mitigate the generalizability of our findings. We studied only three procedures; results might differ for other health care services. We studied common procedures, however, and did not select ones generally perceived to be of no medical benefit. On the contrary, we selected procedures generally agreed— at least under some circumstances—to be beneficial. We studied procedures performed by a variety of physicians as well as diagnostic and therapeutic procedures.

Another factor that might limit the applicability of our findings is that we studied practice in 1981. Medicine has changed since then, and an appropriateness study of current practice might yield different results. While some important changes have occurred—eg, the advent of coronary angioplasty—we do not believe the bulk of community-based practice, which is the subject of this study, has changed sufficiently to render these results outdated.

While detailed data on rates of use, comparable with the data we published previously, are not routinely available, the existing data suggest that the use of the procedures we studied has increased since 1981. For example, the National Center for Health Statistics reported an increase of 34% in the overall rate of use of cardiac catheterization per 100 000 population between 1981 and 1984.³³³ Similarly, the rate of open heart surgery increased by 41% during the same period. In 1984, the same data source reported a national rate of coronary artery bypass surgery of 25 per 10 000 population for persons aged 65 years or older. This rate is greater than the rate we reported for our highest-use site in 1981 (23 per 10 000). We believe that, in the main, the way these procedures are used in the community has changed little since 1981. We know of no evidence to suggest that the appropriateness of use of any of these procedures has improved between 1981 and 1987.

A third limiting factor is that we studied the use of these procedures only among elderly Medicare beneficiaries. Might the results have been different in a younger population? We have no reason to believe so. If a bias exists in studying the elderly, it may be in the direction of revealing even less inappropriate use than the same study of a younger population. To the extent that physicians may tend to be somewhat less aggressive with these procedures in the elderly, we might find even more inappropriate use in a younger group.

We believe these results may have important implications for current health care policy. Our findings should be replicated for other procedures, for populations who are not elderly, and for more recent periods. A consistent finding of significant inappropriate use would challenge us to find ways of selectively eliminating these practices as a method of substantially improving the quality of care we provide and perhaps simultaneously controlling costs.

This work was supported by the Commonwealth Fund, New York, the John A. Hartford Foundation, New York, the Health Care Financing Administration of the US Department of Health and Human Services, Washington, DC, the Pew Memorial Trust, Philadelphia, and The Robert Wood Johnson Foundation, Princeton, NJ.

The authors are deeply grateful to all of the physicians who participated in this study and to the efforts of their staffs who facilitated its work. The efforts of the medical records staffs of the participating hospitals were also indispensable. The study could not have been accomplished without the tireless endeavors of the many data collectors who provided the information on which these analyses depend.

We also wish to thank Mary-Frances Flynn, MPH, Lois McCluskey, MPH, Barbara J. Genesee, and Carol D. Kolven, MA, for their perseverance in overseeing the completion of the fieldwork. Mary Stout and Elizabeth Sullivan deserve special thanks for their talent and commitment in preparing all our manuscripts.

We are also indebted to the members of our Policy Advisory Board whose contributions throughout the study were vital to its success: Robert J. Adolph, MD, John D. Alexander, MD, Richard Berman, MHS, Gerald E. Bisbee, PhD, Ernie Chaney, MD, John H. Dawson, Jr, MD, Linda K. Demko, PhD, Neil J. Elgee, MD, Ronald G. Evans, MD, Donald S. Frederickson, MD, Paul F. Griner, MD, Anthony M. Imparato, MD, John M. Lowe, Douglas B. McGill, MD, Francis D. Moore, MD, Marvin J. Shapiro, MD, Thomas L. Stern, MD, and Richard S. Wilbur, MD.

References

5. Roos NF, Roos LL Jr. Surgical rate variations: Do they reflect the health or socioeconomic characteristics of the population? Med Care 1982;20:945-958.