Appropriateness of Acute Medical Care for the Elderly

An Analysis of the Literature

Robert H. Brook, Caren J. Kamberg, Allison Mayer-Oakes, Mark H. Beers, Kristiana Raube, Andrea Steiner
The research described in this report was supported by grants from the American Association of Retired Persons, the John A. Hartford Foundation, the Robert Wood Johnson Foundation, and by The RAND Corporation in accordance with its program of public service.

ISBN: 0-8330-1010-7

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.

Published by The RAND Corporation
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406-2138
Appropriateness of Acute Medical Care for the Elderly

An Analysis of the Literature

Robert H. Brook, Caren J. Kamberg, Allison Mayer-Oakes, Mark H. Beers, Kristiana Raube, Andrea Steiner

September 1989

Supported by the American Association of Retired Persons
John A. Hartford Foundation
Robert Wood Johnson Foundation

RAND
PREFACE

This work, supported by the American Association of Retired Persons, the John A. Hartford Foundation, the Robert Wood Johnson Foundation, and The RAND Corporation, was undertaken to determine the quality of our knowledge base concerning the appropriateness of care given to the elderly. It is hoped that understanding what is known about appropriateness will be useful to those who are responsible for delivering care to the elderly, those who influence the policies by which care is rendered, and those who are in a position to enhance and extend the knowledge base.
SUMMARY

During the past 30 years, health care expenditures have virtually exploded. Before 1960, health care accounted for 4.4 percent of the gross national product (GNP) (Fuchs, 1986). Today, health care represents over 11 percent of the GNP.

The increase in the over-65 population and the rapid rise in health care costs have combined to create pressures on the health care system. Many have suggested that the way to offset increased expenditures is simply to reduce the amount of care provided, i.e., to ration services. We suggest that, before we even begin to discuss the rationing of care, we must determine whether the care that we are currently paying for is appropriate to the needs and even the wants of the elderly. Is there overuse or underuse of health care? Are the elderly receiving care that will maximize their ability to function? Will reductions in inappropriate use result in substantial savings? Can a clinical-epidemiologic model be developed and implemented that will eliminate most inappropriate care while minimizing a decrease in appropriate care?

This report analyzes the literature regarding appropriateness of acute care provided to the elderly and, using this analysis, suggests policy options that could promote more appropriate care. We discuss three frameworks that can be used to measure appropriateness, and use these approaches to categorize the literature. We present the methods we used to survey and evaluate the appropriateness literature and the results of our literature review. We examine the appropriateness of three kinds of care used by the elderly: hospital, physician, and pharmaceutical. Finally, supported by our findings, we make some research, clinical, and policy recommendations.

In measuring appropriateness, we have delineated three approaches by which to judge whether a service is appropriate. The first approach weighs the risks and benefits of a service. According to this approach, a service is judged to be appropriate only if the benefit of the medical action exceeds the risk of the action by a wide enough margin that the procedure is worth doing. Using the second, or benefit-cost approach, a procedure is judged appropriate when its marginal benefits exceed its marginal costs. The final approach to measuring appropriateness is what we have called the “implicit approach.” Here, standards for judging appropriateness may have been used, but no clear, explicit definitions are specified.
To determine which and how often these three approaches have been used in studies of appropriateness, we conducted a MEDLINE search covering the years 1980 to 1988 to gather those articles that might address the subject. Interestingly, “appropriateness” is not a major heading in the MEDLINE directory and we were thus obliged to take an indirect path to identify articles. As a result, we reviewed articles on quality of care, access, utilization, evaluation, overuse, misuse, underuse, geographic variations, unnecessary care, and geriatric assessment. To broaden our understanding of appropriateness, we also conducted informal interviews with 12 prominent health services researchers around the country.

We reviewed each article and judged the study it described in terms of: (a) whether it used an explicit or implicit definition of appropriateness, (b) whether it explicitly addressed appropriateness in terms of the elderly, (c) its reliability, (d) its internal validity, and (e) its external validity. We identified 10 criteria for evaluation and then created a scoring system where 0 points were given if the criterion was not met, and 1 to 3 points were given if the criterion was “acceptably” or “optimally” met. Articles were evaluated on a 0 to 14 point scale.

We identified 17 articles that explicitly cited appropriate or inappropriate care (including under-, over-, and misuse) provided in hospital and ambulatory settings and for procedures. Ten articles examined the appropriateness of procedures (mean methodology score, 8; range from 2 to 14); six articles looked at the appropriateness of hospital use (mean score, 8; range from 3 to 12); and one article examined the appropriateness of ambulatory care (score 7). Over half of these articles used an implicit approach in defining the appropriateness of care. Although the majority of articles described inappropriate overuse, three articles examined the inappropriate underuse of health care services, and one examined the misuse of hospital services for breast cancer patients.

We identified 19 articles that presented data on the appropriateness of medication use in the elderly population. (Although many articles in the drug literature have titles that imply an examination of appropriateness, very few actually present relevant data on the appropriateness of medication use.) The mean method's score for these studies was 6.2 (range 3 to 11). Eleven of the 19 studies examined overuse of medication.

When we began this work to describe what is known about the level of the appropriateness of acute care rendered to people 65 years of age and older, we knew that the literature was sparse. However, we were not adequately prepared for both the lack of information and the relatively low quality of the studies that we found. For the most part,
information was either out of date or not generalizable. Thus, the first question that faces policymakers, patients, and physicians is whether one wants valid and generalizable information about the appropriateness of the acute services rendered to the elderly in the U.S. health care system. If the answer to that question is yes, it will require the annual investment of tens of millions of dollars.

Even given the state of the literature, it is possible to comment, for at least a few procedures, about the level of inappropriate acute care delivered in the United States. It is reasonable to state that virtually every study included in this review found at least double-digit levels of inappropriate care. Studies that looked at overuse in the hospital—of procedures or of medications—found evidence of such overuse in amounts that are too large to be ignored. In particular, perhaps as much as one-fifth to one-quarter of acute care hospital services or procedures were felt to be used for equivocal or inappropriate reasons, and two-fifths to one-half of the medications studied were overused in outpatients.

The few studies that examined underuse or misuse of services also documented these phenomena. This was especially true for the ambulatory care of chronic physical and mental conditions and concerned the use of low-cost ticket technologies (visits, preventive services, some medications). Thus, we conclude that there appears to be a substantial problem in the matching of acute care services to the needs of elderly patients. This mismatch occurs in terms of both overuse and underuse, at least for areas where research has been conducted.

Our conclusion can be viewed pessimistically as an indictment of the current U.S. health care system or optimistically as a window of opportunity. By recognizing that a large percentage of what we currently do is not needed, it is possible to identify scenarios that would free up resources that could be used to provide better acute care as well as additional resources for long-term care and mental health care to the elderly. Such a strategy, by improving the quality of acute care, might also reduce the prevalence of iatrogenic illness. In addition, experiments at a clinical or patient-doctor level that are designed to improve appropriateness of care can begin now.
ACKNOWLEDGMENTS

The authors thank George A. Goldberg, G. Terry Hammons, and David H. Solomon for their insightful reviews of this report. In addition, we would like to express our thanks to Elizabeth Sullivan for her efforts in typing the several drafts of this report.
CONTENTS

PREFACE ........................................ iii
SUMMARY ........................................ v
ACKNOWLEDGMENTS ............................. ix
TABLES ........................................... xiii

Section
I. INTRODUCTION .............................. 1
II. DEFINING AND MEASURING APPROPRIATENESS ... 3
   Benefit-Risk Approach ...................... 4
   Benefit-Cost Approach .................... 4
   Implicit Approach ......................... 5
III. REVIEW OF THE LITERATURE ............. 6
   Methods ..................................... 6
   Standards for Assessment .................. 7
   Assessing Appropriateness ............... 11
   Procedures, Hospital Stays, and Office Visits 12
   Medications ................................ 20
IV. DISCUSSION .................................. 25

Appendix
A. APPROPRIATENESS OF USE OF PROCEDURES,
   HOSPITALS, AND AMBULATORY CARE:
   AN ANNOTATED BIBLIOGRAPHY .......... 29
B. APPROPRIATENESS OF USE OF MEDICATIONS:
   AN ANNOTATED BIBLIOGRAPHY .......... 40

BIBLIOGRAPHY .................................. 53
TABLES

1. Scoring system for rating articles on appropriateness . . . . 8
2. Appropriateness of procedures, hospital use, and
   office care ....................................................... 14
3. Appropriateness of medication use by medication class . . . 22
4. Appropriateness and expenditures for medications by
   medication class .............................................. 24
I. INTRODUCTION

During the past 30 years, health care expenditures have virtually exploded. Before 1960, health care accounted for 4.4 percent of the gross national product (GNP) (Fuchs, 1986). Today, health care represents over 11 percent of the GNP. Since 1970, the percent of the total federal health budget spent on Medicare, the largest federally funded health program, has increased by one-third (from 45 to 60 percent) (U.S. Bureau of the Census, 1986). In 1987 alone, Medicare paid out over $75 billion (Economic Report of the President, 1988).

Part, although by no means all, of the increase in Medicare outlays can be attributed to demographic changes. Between 1970 and 1984 the total population 65 and older increased from 9.8 to 11.8 percent. At the same time, the 85 and over group had become the fastest growing segment of the elderly population. In 1970, those 85 and older represented 6.6 percent of those 65 and older; in 1984 they represented 9.4 percent (NCHS, 1984).

These demographic trends are expected to continue. When they are combined with increases in the intensity (sophistication) and volume (number) of services being delivered to elderly persons, health care is projected to consume $640 billion by 1990. Of this, $254 billion is expected to be spent on hospitals, $132 billion on physicians' bills, and $44 billion on drugs. Total Medicare costs to the federal government are projected to be $119 billion (Arnett et al., 1986).

Many have suggested that the way to offset the pressures on the health care system resulting from these increased expenditures is simply to reduce the amount of care provided, i.e., to ration services (Evans, 1983; Strauss et al., 1986). Some maintain that the consequence of containing costs in the United States would be rationing such as that found in the British system (Aaron and Schwartz, 1984). Others have suggested ways to limit our health care spending by discontinuing reimbursement for many forms of acute care for the very old (Callahan, 1987).

Some changes designed to reduce health care costs have already occurred. For example, the way hospitals are reimbursed has been altered, resulting in a prospective payment system for Medicare that is based upon diagnostic related groups (DRGs). In addition, some states have eliminated health insurance (medical coverage) for medically indigent adults, increased coinsurance or deductibles in their Medicaid programs, or encouraged capitated pre-payment systems. Some of these
economic solutions have been found to cause negative effects on health, especially for the sick and disadvantaged (Brook et al., 1983; Lurie et al., 1984; Ware et al., 1986; Soumerai et al., 1987).

We suggest that, before we even begin to discuss the rationing of care according to one or more of the economic solutions noted above, or to condemn or celebrate the effects of prospective payment or capitation, we must determine whether the care we are currently paying for is appropriate to the needs and wants of the elderly. Is there overuse or underuse of health care? Are the elderly receiving care that will maximize their ability to function? Will reductions in inappropriate use result in substantial savings? Can a clinical-epidemiologic model be developed and implemented that will eliminate most inappropriate care while minimizing a decrease in appropriate care (e.g., one that provides better information to the patient and doctor about the circumstances under which the use of a specific procedure improves health)?

If we could explicitly define and measure appropriateness, and if we could specify what constitutes appropriate care for each patient, we would know the best set of procedures or treatments to use for a particular patient with a particular illness. We would know which services Medicare and private insurers should cover and when. We would know when patients should be permitted or denied access to particular services, or at least under which circumstances we should subsidize care with pretax or public dollars. We would know when and how to consider patient preferences. We would know in which setting (e.g., hospital or physician's office) care should be provided. In this “ideal” system, resources would be distributed appropriately among the various components of medical care (e.g., hospitals, home health care agencies, nursing homes).

This report reviews the literature that examines the appropriateness of acute care provided to the elderly and, using this literature review, suggests options at both the policy and clinical levels that could promote more appropriate care. The next section describes our attempt to define and measure appropriateness. We discuss three frameworks by which appropriateness can be judged and use these approaches to categorize the literature. In the following section we present the methods we used to survey and evaluate the appropriateness literature and then present the results of our literature review. We examine the appropriateness of three kinds of care used by the elderly: hospital, ambulatory care, and pharmaceutical. Finally, using our findings, we discuss some research and policy recommendations.
II. DEFINING AND MEASURING APPROPRIATENESS

A service is provided appropriately when its benefits (expected positive consequences) exceed its costs (expected negative consequences). This is a simple definition, but it is not easy to apply. Application requires deciding exactly what is to be considered a benefit and what is a cost, and then determining how to measure them. Measurement is complicated by the need to compare disparate consequences such as longer life, pain, and monetary cost.

A full utilitarian approach would have to account for all benefits and all costs, no matter what they are and no matter to whom they accrue. The benefits to the patient may include increased life span, decreased morbidity, improved function, and decreased worry about health. The costs to the patient may include the risk of a bad outcome from the service, the pain and discomfort of undergoing the service, the time spent, and the portion of the money cost of the service that is not reimbursed by insurance or other outside sources.

Additional benefits and costs may accrue to the patient’s family and friends, to health care providers, and to the remainder of society. Family and friends may benefit through longer and higher quality companionship from the patient if the service is successful, from their share of larger economic earnings by the patient, and because they benefit altruistically from anything that benefits the patient. They may also bear part of the time and money costs of the service and may worry about the risks and discomforts suffered by the patient. Providers may benefit because they are paid to perform the service. They bear as a cost the forgone opportunity to do something else productive or pleasurable with the time spent on providing the service. Others in society may also accrue benefits and bear costs of the service, even though they do not personally know the patient and are not involved in his treatment. The benefits of successful treatment include the additional productive contributions to the economy that the patient may then make. The costs include insurance or tax copayments for the portion of the service not paid for by the patient or his family.

Obviously, the full utilitarian approach cannot be routinely applied, and practical attempts to define and measure appropriateness have used one or both of two simplifications: (1) restriction of the scope of the benefits and risks considered, or (2) use of implicit evaluation instead of actual measurement of those benefits and costs.
We have delineated three practical approaches by which to judge whether a service is appropriate. We identified these approaches through our literature review and interviews with experts in health services research. We have labeled the first two—the benefit-risk and benefit-cost approaches—as "explicit," because they explicitly specify which costs and benefits are to be considered in determining whether a service is appropriate. (Note, however, that even the "explicit" methods typically rely on implicit judgments about the magnitude of these costs and benefits.) The third approach relies entirely on implicit judgments, without specifying what the appropriateness rater should take into account or what he should disregard.

BENEFIT-RISK APPROACH

The first approach considers only the benefits and costs of the service to the patient. According to this approach, a service is judged to be appropriate only if the benefits exceed the negative consequences to the patient by a wide enough margin that the procedure is worth doing. Benefits include increased life expectancy, relief from pain, reduction in anxiety, and improved functional capacity. Negative consequences ("risks") include morbidity, mortality, work loss, and anxiety in anticipation of the procedure. All monetary or resource costs of performing the service are explicitly excluded. Thus, in this approach, factors such as whether a person can afford the procedure, has insurance for it, or would rather have had the money to buy a house are not included. Under the benefit-risk approach, a service that provides a small net benefit to the patient would be considered appropriate, no matter how expensive it is to provide that service. This approach was used in the Health Services Utilization Study (HSUS) to examine the reasons for geographic variations in health services (Chassin et al., 1986, 1987a, 1987b; Park et al., 1986).

BENEFIT-COST APPROACH

The benefit-cost approach extends the benefit-risk approach by considering the resource cost of providing the service, without regard to who bears that cost. The resource cost of the service is in practice typically approximated by the billed cost, whether the bill is paid by the patient, his insurance carrier, the health care provider, or the taxpayer. For a service to be judged appropriate, the net benefits to the patient (the same net benefits that are evaluated under the benefit-risk approach) must exceed the resource cost of the service. If the cost
exceeds the benefits, then society is paying too much for the service, and society should use the resources to provide the service to another patient who would benefit more, or to pay for other health care activities within or outside the medical sector (e.g., long-term care).

IMPLICIT APPROACH

The final approach to measuring appropriateness is what we have called the "implicit approach." Here, standards for judging appropriateness may have been used, but no clear, explicit definitions are specified. Traditionally used in hospital utilization review, implicit approaches are often applied to determine whether or not a patient needs to be admitted to a hospital or needs skilled nursing home care. In general, the implicit approach relies upon the physician to judge appropriateness. The physician, in turn, draws upon both training and clinical expertise to determine whether a specific procedure, hospitalization, or medication was appropriate. However, it is unclear what appropriateness model the physician has in mind when he makes such a judgment.

The three approaches will ordinarily lead to different conclusions about appropriateness. All services judged appropriate by the benefit-cost approach will also be appropriate when judged by the benefit-risk approach, but not vice versa. From the patient's viewpoint, if he has full insurance, he would want all services judged appropriate by the benefit-risk approach because he would then be receiving the maximum total benefit from health care. In contrast, society as a whole might prefer the benefit-cost approach because services would then be provided only if their benefits (to the patient) exceeded their resource cost.

The implicit approach can in theory be either more or less restrictive than the explicit approaches, depending on what factors the rater considers and what relative importance he attaches to them. Implicit judgments based on strong beliefs supporting the application of medical technology will likely classify some services appropriate that are inappropriate according to the explicit approaches, sanctioning some services even when they provide no net benefit to the patient. The same result could follow if the implicit judgment put inordinate weight on the monetary benefits to the physician of providing the service.
III. REVIEW OF THE LITERATURE

METHODS

To determine which and how often these three approaches have been used in studies of appropriateness, we conducted a MEDLINE search covering the years 1980 to 1988 to gather those articles that might address the subject. Literature reviews from prior years were not examined because their data would have been over a decade old and because the appropriateness of care among the elderly was not much investigated before about 1980. Our initial search identified over 300 articles. We identified others by talking to experts, examining reference lists, and scanning the tables of contents of health services research journals. This report is limited to articles that provide quantitative information about the appropriateness of acute hospitalizations, physician visits, use of procedures, and use of medications. We eliminated articles on the appropriateness of dental care, long-term care, mental health care (except in the case of medications), or social services. We also eliminated articles that contained only utilization data rather than judgments or determinations of appropriateness of use. In addition, we eliminated articles that were not applicable to the U.S. health care system (although studies based on European and Canadian data were included) or were not in English. Finally, with the exception of articles about medications, we excluded studies that did not explicitly include the elderly. We defined “elderly” as anyone over the age of 50. An article on the appropriateness of medication use was included if the elderly are typical users of the specific drug even if the population studied did not include those over 50. For studies on medications that included an intervention aimed at improving appropriate prescribing, we present only the data on the pre-intervention results.

Interestingly, “appropriateness” is not a major heading in the MEDLINE directory, and we were thus obliged to take an indirect path to identify articles. This resulted in reviewing articles on quality of care, access, utilization, evaluation, overuse, misuse, underuse, geographic variations, unnecessary care, and geriatric assessment. Sometimes the articles we reviewed did not use the word appropriateness explicitly, but rather addressed the subject within the context of our three definitions and were therefore included. Other articles used the word appropriateness but on review did not address the subject and were therefore excluded.
To broaden our understanding of appropriateness we also conducted informal interviews with 12 prominent health services researchers around the country. These included physicians, economists, policy analysts, geriatricians, and social scientists from RAND, UCLA, Harvard University, Boston University, and the Veteran's Administration.

What follows is the synthesis of the literature and interview processes. We first present the criteria and explain how we rated the methods of the studies we evaluated. This is followed by a detailed examination of the articles we identified that address the appropriateness of doctor visits, procedures, acute hospitalizations, and medications. The last section summarizes the findings and makes policy and research recommendations.

STANDARDS FOR ASSESSMENT

After eliminating articles based on the criteria discussed above, we reviewed each remaining article and judged the study it described in terms of: (a) whether it used an explicit or implicit definition of appropriateness, (b) whether it explicitly addressed appropriateness in terms of the elderly, (c) its reliability, (d) its internal validity, and (e) its external validity. We identified 10 criteria for evaluation and then created a scoring system where 0 points were given if the criterion was not met at all, and 1 to 3 points were given depending on the degree to which the criterion was met (Table 1). This resulted in a scale of possible scores from 0 to 14 points. The weighting system used was based on the belief that all of the dimensions were of relatively equal importance. The methodology was not subjected to any formal testing.

As an example of the scoring for an individual article, a study received a total of 8 points when it used a benefit-risk approach (1 point), used explicit criteria to develop standards used to judge appropriateness (1 point), presented the sample breakdown in terms of those over and under age 65 (1 point), tested the reliability of the data collection method (1 point) and the judges (1 point) but did not provide the actual reliability results (0 points), demonstrated face validity (1 point), had a sample size of 100 or more (1 point), was conducted at a single institution (0 points), and demonstrated representativeness in terms of its sample’s demographics (1 point). Actual scores were determined by three independent raters. In the case of disagreements, differences were discussed and articles were read and scored again. Final results were averages of the three raters’ scores. We assigned a rating (average, above average, below average) to each appropriateness article we reviewed based on the mean score in that article’s category.
Table 1
SCORING SYSTEM FOR RATING ARTICLES ON APPROPRIATENESS

<table>
<thead>
<tr>
<th>Characteristics of the Approach Used to Measure Appropriateness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Implicit vs. Explicit Approach</td>
</tr>
<tr>
<td>0 = implicit approach</td>
</tr>
<tr>
<td>1 = benefit-risk or benefit-cost approach</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type of Standards Used to Judge Appropriateness</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = implicit criteria: criteria not explained (e.g., based only on author’s subjective opinion)</td>
</tr>
<tr>
<td>1 = explicit criteria used</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analysis by Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Addressed Separately or as Part of Larger Group</td>
</tr>
<tr>
<td>0 = sample includes people over age 50, but results are not presented by age category</td>
</tr>
<tr>
<td>1 = analysis presented by age category, 65 and over</td>
</tr>
<tr>
<td>2 = analysis presented by subcategories of age 65 and older (e.g., 65–74, 75–84)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Reliability of Data Collection</td>
</tr>
<tr>
<td>0 = not tested</td>
</tr>
<tr>
<td>1 = tested</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Reliability of Judges</td>
</tr>
<tr>
<td>0 = not tested</td>
</tr>
<tr>
<td>1 = tested</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Level of Reliability Met</td>
</tr>
<tr>
<td>0 = not reliable or results not given</td>
</tr>
<tr>
<td>1 = if researchers conducted reliability tests (as described in the text), they reported reliability for either data collection or judges' consensus panels</td>
</tr>
<tr>
<td>2 = if researchers conducted reliability tests (as described in the text), they reported reliability for both data collection and judges' consensus panels</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Internal Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. Judgment Processes</td>
</tr>
<tr>
<td>0 = if explicit, criteria lack face validity; if implicit, author is only judge</td>
</tr>
<tr>
<td>1 = if explicit, the criteria had face validity; if implicit, judges selected from a defined sample frame of experts nominated by an accredited professional body</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>External Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Sample Size</td>
</tr>
<tr>
<td>0 = under 100</td>
</tr>
<tr>
<td>1 = 100 or more</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nature of Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>9. Number/Nature of Sites</td>
</tr>
<tr>
<td>0 = single institution</td>
</tr>
<tr>
<td>1 = multiple institutions</td>
</tr>
<tr>
<td>2 = single geographical area or enrolled population (e.g., HMO)</td>
</tr>
<tr>
<td>3 = multiple geographical areas or enrolled populations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Representativeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. Representativeness in Terms of Age, Sex, Race, and Income</td>
</tr>
<tr>
<td>0 = not representative</td>
</tr>
<tr>
<td>1 = representative</td>
</tr>
</tbody>
</table>

---

*These terms are defined in detail in the text.*
of care (i.e., procedures, ambulatory visits, acute hospitalizations, medications). An article was considered average if its score was plus or minus 2 points from the mean.

In reading Table 1, the following definitions of reliability and validity should be kept in mind.

*Reliability* makes a statement about the extent to which measures are repeatable (Nunnally, 1978). Tests of reliability can be made by comparing results from alternative forms of the same test, by duplicate tests given at different times but under identical circumstances (test-retest reliability), or by using different examiners (inter-rater reliability). If the data are influenced strongly by random measurement error, the results will not be repeatable. In statistical terms, reliability is the proportion (from 0 to 100 percent) of total measured variance that is real.

In its simplest terms, *validity* refers to the extent to which a measure assesses what it purports to measure. Researchers describe two categories of validity: internal and external.

*Internal validity* addresses those factors that guarantee that the study's results are "true" (i.e., whether a study's results are due to the planned intervention or to external factors). In other words, is the research defensible in terms of its design, data collection methods, and analytic plan? Face validity refers to the extent to which an instrument "looks like" what it is intended to measure. The term basically arises from questioning whether a standard makes sense "on the face of it" (Turner and Martin, 1984; Nunnally, 1978).

*External validity* refers to those aspects of a research design that allow it to be generalized and applied in a broader context. Such design features as method of sample selection, sample size, representativeness, and geographical diversity contribute to (or detract from) external validity. For example, research that examines appropriate health care among those 40 years of age and younger may not be able to be generalized to a population that includes those age 65 and over.

Finally, a sample is said to be *representative* if it imitates the population under consideration, i.e., if it is a miniature version of that population in terms of demographics such as age, sex, race, and income (Fink and Kosecoff, 1978). For example, if a disease affects both genders, both are studied (e.g., lung cancer); if it affects more than one race, at least two racial groups are included; if it affects more than one age group within the elderly (e.g., 65–74, 75 and older), both are studied; and if the disease affects persons in more than one income group, all are included.

Although not included in our scoring system, an important component of a study's methods is how the data were collected.
Medical Record Review

Audits of hospital and physician offices' patient medical records provide a rich supply of information on both the process (what was done to patients) and outcome (what happens to patients) of care. These data can be used to measure appropriateness. One major drawback of this data source, however, is that records may be incomplete.

Patient or Physician Interviews

Using interviews, patients have been asked to describe their health in detail, to describe what their doctor did, and to assess their satisfaction with their health care. Information of this type can be used to assess appropriateness. Researchers have, however, debated medically uneducated patients' ability to accurately assess the quality or appropriateness of the care they receive. Some are persuaded that patients can be very good reporters of information (Davies and Ware, 1988), whereas others maintain they cannot (Kisch and Reeder, 1969).

In addition, physicians can be asked why they performed a procedure. Major drawbacks of such a method are that it is sometimes difficult to locate the physician who made this decision and that the physician might answer questions so as to put him or her in a favorable light.

Claims Data

Finally, researchers may use insurance claims files as a way to assess appropriateness. This is particularly effective for assessing appropriateness in elderly patients because virtually all older Americans are Medicare-eligible. An important limitation is that insurance files lack the detailed clinical data that are required to assess appropriateness. In addition, information about services not covered by the insurance plan (e.g., medications for Medicare patients) is not contained in these databases. However, linking eligibility files and insurance claims data to detailed clinical data obtained from inpatient or outpatient medical records will result in a data base that is both useful for assessing appropriateness of care and is generalizable in its conclusions.
ASSESSING APPROPRIATENESS

In the preceding sections, we have proposed frameworks for assessing appropriateness, showed how some researchers have assessed them, and set standards for evaluating the literature. In this section, we discuss the appropriateness of procedures, hospital stays, office visits, and medications. We examine the definitions that have guided research in this area, review the methods used, describe the available data, evaluate the literature, and discuss the need for further research. We have concentrated on articles that assess appropriateness of use for the elderly. In the case of medications, because many of the data on appropriateness do not exclusively examine use of medications prescribed for the elderly, we have included all information that does not expressly exclude the elderly or that examines the kinds of medications that would be used in elderly populations.

The literature suggests that elderly populations demonstrate underuse, misuse, and overuse of health care services (Rice and Feldman, 1983). The underuse of a procedure, office visit, or hospital stay is the failure to use these services when they are indicated. The underuse of a medication is the failure to use a medication if its use is indicated, whether the indication is for the specific medication or for an equally efficacious one. For example, under most circumstances, the failure to vaccinate elderly persons for influenza is inappropriate (Baker and Mullowsky, 1980). Regardless of the medication considered, the failure to treat a patient with angina is also inappropriate.

Misuse implies that services are incorrectly used. For example, it may be appropriate to admit an elderly patient with pneumonia to the hospital, but inappropriate to discharge him to home within 48 hours when he is still too weak to get out of bed. Thus, duration of hospital stay, either too long or too short, might be considered misuse. Similarly, the misuse of medication applies to the incorrect method of using a medication although its use is otherwise indicated. This may include errors of dose, frequency, route, duration, or monitoring. For example, beginning tricyclic medication at high dose in the elderly is misuse (Veith, 1982).

Overuse is the use of a procedure, office visit, or hospital admission that is not indicated. For example, the overuse of a medication occurs when that specific medication should not be used or when no medication at all should be used. For example, the use of antibiotics in uncomplicated upper respiratory infections is inappropriate (Wyngaarden and Smith, 1985), regardless of which antibiotic is used. When tricyclic antidepressants are used in an elderly person to treat insomnia rather than depression, it constitutes overuse (Duncan and Campbell, 1988).
PROCEDURES, HOSPITAL STAYS, AND OFFICE VISITS

Review of the Literature

We identified 17 articles that explicitly cited appropriate or inappropriate care (including under-, over-, and misuse) provided in hospital and ambulatory settings and for procedures. Appendix A details each article’s methods. Ten articles examined the appropriateness of procedures (mean methodology score, 8; range from 2 to 14); six articles looked at the appropriateness of hospital use (mean score, 8; range from 3 to 12); and one article examined the appropriateness of ambulatory care (score 7) (Table 2).

Over half of these articles used an implicit framework in measuring the appropriateness of care. None used the benefit-cost approach. Although the majority of articles described inappropriate overuse, three articles examined the inappropriate underuse of health care services (Melton et al., 1982; Greenfield et al., 1987; Heller et al., 1984), and one examined the both misuse and underuse of hospital services for breast cancer patients (Greenfield et al., 1987).

Procedures—Background

The number of surgical procedures received per year by the elderly has been increasing since the inception of Medicare. In 1965, the number of operations per 1,000 aged persons was 105 (Aging America, 1984); in 1973 it was 141 per 1,000 aged persons (NCHS, 1976); and in 1986 it was 409, a 290 percent increase over 1965 (NCHS, 1988a).

Procedures— Appropriateness

Table 2 provides information on the methods, e.g., sample size, population source, and results of the 10 studies we reviewed on the appropriateness of procedures. One of the studies used a sample that came from only one site and five used an implicit definition of appropriateness. All but one of the articles described inappropriate overuse of procedures or tests and almost all found double-digit levels of inappropriateness. Although the data in Table 2 cannot be averaged, the general impression from them is that a considerable amount of overuse exists. Nevertheless, little is known about the appropriateness of the use of procedures in the elderly. Some specific highlights follow.

Only a few studies examined the appropriateness of care within specific age subgroups. Winslow and her RAND colleagues (Winslow et al., 1988a) analyzed the appropriateness of coronary artery bypass
surgery for Medicare patients in one western state. They found that 14 percent of these procedures were performed for clearly inappropriate reasons and that 30 percent of these procedures were performed for equivocal reasons, i.e., use not clearly appropriate or inappropriate. Appropriateness of use for patients 65 years of age and older was at least as high as that for younger groups. There was, however, considerable variation by the hospital in which the procedure was performed, ranging from 6 to 23 percent for inappropriate reasons and 17 to 40 percent for equivocal reasons. The RAND researchers also found a high level of inappropriate overuse of coronary angiography, endoscopy, and carotid endarterectomy in three large geographic areas in 1981 (Chassin et al., 1986, 1987a, 1987b; Winslow et al., 1988b). If equivocal use is combined with inappropriate use, then 26, 28, and 64 percent, respectively, of angiography, endoscopy, and endarterectomy was not justified.

Regarding the level of appropriateness of use for coronary angiography, large discrepancies were found between the RAND study and Elliott et al.'s (1981) study. Specifically, the RAND group found that 17 percent of coronary angiography was inappropriate, whereas Elliott's group found that less than 3 percent was inappropriate. The explanation for this large difference may lie in the reliability and validity of the two studies' methods and the appropriateness judgment process (see Appendix A). For instance, Elliott et al. specified five indications for performing the procedure and six for not performing it. These indications were much more general than the 300 RAND indications.

The hip arthroplasty study represents an indirect attempt to define underuse. Although its conclusions are mostly speculative, it is included in our review because it demonstrates the need to begin work in this area. This study began by assuming that all hip replacements in one Minnesota county served by the Mayo Clinic were appropriate (Melton et al., 1982). No explicit appropriateness criteria were developed, nor was an implicit review performed to support this assertion. The Olmstead County rate was applied nationally after adjustments for age, sex, and race, and a need for 40 percent more hip operations was found. The findings from this study are interesting but are not based on any direct measurement of underuse.

The value to the policy process of results from studies of appropriateness would become more apparent if data from them could be linked to information on the amount of resources consumed by such procedures. Even though we searched for such information, except for a few procedures, this unfortunately cannot be done because of limited data on appropriateness.
<table>
<thead>
<tr>
<th>Health Care Service</th>
<th>Population Size</th>
<th>Population Source</th>
<th>Approach to Measurement</th>
<th>Type of Inappropriateness</th>
<th>Percentage of Inappropriateness</th>
<th>Quality Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronary artery bypass surgery</td>
<td>1. 386</td>
<td>1 state</td>
<td>Benefit-risk</td>
<td>Overuse</td>
<td>14%</td>
<td>14</td>
</tr>
<tr>
<td>Coronary angiography</td>
<td>2. 1,677</td>
<td>3 regions</td>
<td>Benefit-risk</td>
<td>Overuse</td>
<td>17%</td>
<td>13</td>
</tr>
<tr>
<td>Carotid endarterectomy</td>
<td>3. 1,302</td>
<td>3 regions</td>
<td>Benefit-risk</td>
<td>Overuse</td>
<td>32%</td>
<td>13</td>
</tr>
<tr>
<td>Upper GI endoscopy</td>
<td>4. 1,585</td>
<td>3 regions</td>
<td>Benefit-risk</td>
<td>Overuse</td>
<td>17%</td>
<td>13</td>
</tr>
<tr>
<td>Carotid endarterectomy</td>
<td>5. 95</td>
<td>5 hospitals</td>
<td>Benefit-risk</td>
<td>Overuse</td>
<td>13%</td>
<td>9</td>
</tr>
<tr>
<td>Multiple procedures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appendectomy</td>
<td>6. 4,850</td>
<td>1 state</td>
<td>Implicit</td>
<td>Overuse</td>
<td>3%</td>
<td>6</td>
</tr>
<tr>
<td>Cataract removal</td>
<td></td>
<td>Medicare patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cholecystectomy</td>
<td></td>
<td>Medicaid patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coronary angiography</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uterine dilation &amp; curettage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hial hernia repair</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal hysterectomy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaginal hysterectomy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lumbar disk excision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meniscectomy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tonsillectomy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenoidectomy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hip arthroplasty</td>
<td>7. 321</td>
<td>1 county</td>
<td>Implicit</td>
<td>Underuse</td>
<td>40%</td>
<td>4</td>
</tr>
<tr>
<td>Preoperative lab screening</td>
<td>8. 2,000</td>
<td>1 site</td>
<td>Implicit</td>
<td>Overuse</td>
<td>60%</td>
<td>3</td>
</tr>
<tr>
<td>Cardiac pacemaker</td>
<td>9. 382 Medicare patients</td>
<td>30 hospitals</td>
<td>Implicit</td>
<td>Overuse</td>
<td>20%</td>
<td>3</td>
</tr>
<tr>
<td>Cardiac pacemaker</td>
<td>10. Various</td>
<td>Various</td>
<td>Implicit</td>
<td>Overuse</td>
<td>30–75%</td>
<td>2</td>
</tr>
<tr>
<td>Health Care Service</td>
<td>Population Size</td>
<td>Population Source</td>
<td>Approach to Measurement</td>
<td>Type of Inappropriateness</td>
<td>Percentage of Inappropriateness</td>
<td>Quality Score</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>-------------------------</td>
<td>---------------------------</td>
<td>---------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Hospital Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital admission &amp; day of care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. 1,132</td>
<td>6 sites in 4 regions</td>
<td>Benefit-risk</td>
<td>Overuse</td>
<td>10%–35%</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>2. All Medicare &amp; Medicaid</td>
<td>25 hospitals in 4 regions</td>
<td>Implicit</td>
<td>Overuse</td>
<td>12%–28% hospital admissions</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Hospital days of care only</td>
<td>6 hospitals</td>
<td>Implicit</td>
<td>Overuse</td>
<td>7% of hospital admissions or hospital days of care</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>3. 6,574 Medicare</td>
<td>6 hospitals</td>
<td>Benefit-risk</td>
<td>Overuse</td>
<td>33% fee-for-service inappropriate; 27% HMO inappropriate</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>4. 200</td>
<td>1 site</td>
<td>Benefit-risk</td>
<td>Overuse</td>
<td>6%</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Hospital care for breast cancer patients</td>
<td>4 hospitals</td>
<td>Implicit</td>
<td>Overuse</td>
<td>4–17%</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>5. 1,456 Medicare</td>
<td>7 hospitals</td>
<td>Benefit-risk, misuse</td>
<td>Underuse</td>
<td>26–80%</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>6. 374 patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Office Visits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For 5 common conditions</td>
<td>1,226</td>
<td>Population-based</td>
<td>Implicit</td>
<td>Underuse</td>
<td>26–80%</td>
<td>7</td>
</tr>
</tbody>
</table>
Each year, one in 500 Americans (mostly elderly) receive a pacemaker at an approximate annual expense of $2 billion (Greenspan et al., 1988). Greenspan et al. found that 20 percent of pacemaker implantations are inappropriate. Generalizing these results, almost half a billion dollars a year could be saved for this one procedure alone if its inappropriate use were eliminated, and if other medical expenses were not incurred by those people who did not receive a pacemaker.

Another procedure that is performed frequently in the United States is coronary artery bypass surgery. In 1985, an estimated 200,000 procedures were performed in the United States, with costs ranging from $18,000 to $59,000 per procedure (Los Angeles Times, 1988). Winslow states that "few would argue with the proposition that the services should go first to all patients with clearly appropriate reasons (55 percent of all procedures were performed for appropriate reasons) for their use" and "... if society wishes not to pay for procedures in the equivocal or inappropriate categories, then one could almost double the number of appropriate coronary artery bypass surgeries without raising health care expenditures" (Winslow et al., 1988b). Again, using some simple assumptions, perhaps $2.2 billion could be saved if $25,000 were saved on each of 90,000 procedures. Applying the same naive reasoning to carotid endarterectomy and eliminating both equivocal and inappropriate reasons for performing the procedure suggests that $1.0 billion could be saved ($15,000 average cost/procedure; 100,000 procedures performed in recent years; two-thirds inappropriate or equivocal). (Note that if only inappropriate, as opposed to inappropriate and equivocal, coronary artery bypass surgeries and carotid endarterectomies were eliminated, respective savings would be $850 million and $500 million.)

Of course, reducing health care expenditures is by no means the only issue when considering the appropriateness of procedures. For example, many procedures have high complication rates, especially in elderly populations. Winslow et al. (1988a) recently reported on the appropriateness of and complications from carotid endarterectomy, a procedure that has been thought to decrease strokes. Over 107,000 of these procedures were performed in 1985. Winslow et al. found that approximately one-third of the procedures in their study were used inappropriately. In addition, even when the procedure was used appropriately, the average complication rate (i.e., strokes and deaths produced by the procedure) was almost twice the acceptable level. (Acceptable level of complication rates was defined as the level below which the number of strokes and deaths caused by the procedure was lower than would have been expected if the patient did not have the procedure.) Winslow et al. suggested that even procedures performed for appropriate
indications might be unwarranted because, given the unexpectedly high complication rate, the risks of the procedure would outweigh the benefits for many of the indications that the panel, which had assumed lower complication rates, had previously considered appropriate. They concluded that this procedure should be limited to surgeons and institutions that have known low complication rates.

**Hospital Use—Background**

Expenditures on hospital care account for over two-thirds of the dollars Medicare spends for personal health care. The total per capita amount spent on hospital care for persons age 65 and over was $1,900 in 1984 (Aging America, 1988). Since Medicare began in 1965, discharge rates from short-stay hospitals for persons age 65 and over have increased by 39 percent, from 264 per 1,000 persons in 1965 to 367 per 1,000 persons in 1986 (NCHS, 1970, 1988b). (Rates decreased from 389 to 367 per 1,000 between 1979 and 1986 (NCHS, 1981, 1988b).)

**Hospital Use—Appropriateness**

We identified six articles that described the appropriateness of hospital use. All of them included more than one hospital, three used an explicit definition of appropriateness, and all but one examined inappropriate overuse of hospital services (Table 2). The majority of articles used an explicit set of criteria, the Appropriateness Evaluation Protocol (AEP) developed by Gertman and Restuccia, to judge appropriateness (Gertman and Restuccia, 1981; Restuccia et al., 1984; Siu et al., 1986, 1988). The AEP criteria determine the appropriateness of a hospital admission, the appropriateness of each hospital day of care, and the appropriateness of the site of care (e.g., would ambulatory surgery be as appropriate?). The AEP contains 18 possible admission criteria and 27 possible day-of-care criteria. For example, it might be appropriate for an elderly patient with pneumonia to be admitted to the hospital, but it may be inappropriate for that same patient to remain in the hospital for 15 days. If any single admission criterion is met, the admission is considered appropriate; similarly, if any day-of-care criterion is met, the day of care is considered appropriate. An integral feature of the AEP is the allowance of a change of appropriateness judgment based on the subjective opinion of the judges. Thus, hospital care that did not meet the explicit criteria of the AEP could be judged appropriate and conversely hospital care judged appropriate by the AEP could be judged inappropriate based on a subjective override by a physician reviewer.
The articles we reviewed found a substantial amount of hospital overuse. Reducing hospital overuse is important because it could lower the expenditures for the hospitalized elderly as well as potential iatrogenic illness that can accompany hospitalization. One study found that one in five admissions and one in four days of care were inappropriate (Restuccia et al., 1984). Another study found that 22 percent of the patients admitted to the hospital had at least one inappropriate hospital day (Restuccia, 1982), and a third study demonstrated that 23 percent of admissions and 34 percent of hospital days were inappropriate (Siu et al., 1986).

Three studies examined factors that might reduce hospital overuse. Two of them used concurrent physician feedback of the appropriateness judgment as an intervention to reduce inappropriate hospital care (Restuccia et al., 1984; Studnicki and Stevens, 1984). Both studies found that feedback reduced inappropriate hospital days of care. The third study examined how type of health insurance (i.e., fee-for-service, capitated) might affect overuse (Siu et al., 1988).

Health maintenance organizations (HMOs) are known to achieve their cost savings through lower rates of hospital admissions, but little is known about the appropriateness of these lower rates. Siu et al. (1988) found that inappropriate discretionary surgery was lower in an HMO, but that the rate of inappropriate nondiscretionary surgery (i.e., emergency surgery) was the same in an HMO as in a fee-for-service setting. The overall rate of hospital admissions judged to be discretionary was 32 percent in the fee-for-service system, and 27 percent in the HMO. Unlike the previous work using the AEP, which considered an admission appropriate if a service requiring hospitalization was provided, the Siu study used both this criterion as well as whether the treatment requiring hospitalization was medically warranted.

One article found that there may be a significant amount of inappropriate underuse and misuse of surgery for elderly women with breast cancer. Greenfield et al. (1987), in a study of the appropriateness of hospital treatment for breast cancer, found that the appropriateness of breast cancer treatment was related to age. Specifically, they found that 4 percent of patients 30 to 69, compared with 17 percent of patients 70 and over, received too little surgical therapy for breast cancer.

In summary, as with procedures, we know very little about the appropriateness of hospital use, and most of our information comes from a period before prospective payment was implemented. However, if the situation has not changed, the 15 to 30 percent inappropriate rates of hospital use cited in these studies are too high to be ignored.
Ambulatory Care—Background

In 1984, persons aged 45 to 64 averaged 6.6 medical visits per year, compared to 9.1 visits per year for persons ages 65 and older (Aging America, 1988). Since Medicare began, the average number of physician contacts and the percentage of persons 65 and over reporting that they had seen a physician in the last year have both increased substantially, particularly for persons with low incomes. Furthermore, the aging of the population will create a greater demand for physician care in the future. According to projections based on 1980 physician visit rates and U.S. Bureau of the Census population projections, the need for physician care will increase 22 percent by the year 2000 and 125 percent by the year 2030 (Aging America, 1988).

Ambulatory Care—Appropriateness

We found one article that studied the appropriateness of ambulatory care in the elderly (Table 2). Heller et al. (1984) examined the underuse of appropriate services using a population-based sample of elderly aged 65–74 years (the 1975 Health and Nutrition Examination Survey, or HANES). They found a range of 26 to 80 percent “deficient” care for five common medical conditions in the elderly: hypertension, angina, hearing impairment, shortness of breath on exertion, and depression. Hypertension had the least amount of deficient care and depression had the most. In this study, underuse was defined as either physicians underutilizing appropriate treatments or tests for patient complaints such as chest pain or hearing loss, or patients not achieving a specified outcome.

Heller et al. examined whether inappropriate care was due to patient-related factors (e.g., the patient with the symptom did not seek a physician visit), or physician-related factors (e.g., the physician did not evaluate or treat the patient’s problem appropriately). Although both physician- and patient-related factors contributed to deficient care, persons presenting with one of three tracer conditions (shortness of breath on exertion, hearing impairment, depression) had deficient care primarily because of physician-related factors. Thus, although patients sought help for their complaints, physicians did not meet the minimal standards of care as defined by the authors.
MEDICATIONS

Review of the Literature

We identified 19 articles that presented data on the appropriateness of medication use in the elderly (or for drugs commonly used among the elderly even if the specific study did not include older persons). In doing our search, we found that, although many articles in the drug literature have titles that imply an examination of appropriateness, very few actually present relevant data on the appropriateness of medication use.

The mean methodologic score for these 19 studies was 6.2 (range 3 to 11). Three studies were rated as below average, 12 as average, and 4 as above average. A benefit-risk approach was used in three papers and a benefit-cost approach was used in two. In the remaining 14 studies the model of appropriateness was implicit.

Medications—Background

Each year Americans spend over $25 billion on medicines (Baum et al., 1985). Americans over the age of 60 constitute 16 percent of the population but consume 38 percent of all medications, filling over 560 million prescriptions yearly (Baum et al., 1985). At any one time, a person over the age of 65 takes an average of 4.5 medications (2.3 prescription, 2.1 nonprescription) (Ostrom et al., 1985). Several descriptive studies have demonstrated that there are inherent risks in using medications. For instance, nearly 10 percent of acute hospitalizations for the elderly are related to drug toxicity and 17 percent of patients in the hospital develop a toxic reaction to a medication (Soumerai and Avorn, 1984). One out of every 1,000 hospitalized persons followed in the Boston Collaborative Series died as a result of an adverse drug reaction (Porter and Jick, 1977).

Baum et al. (1985), in an analysis of national patterns of medication use by drug class, showed that the community-based elderly (60 and older) use anti-infectives more often than any other class. This use accounts for over 13 percent of total drug expenditures. Expenditures for cardiac drugs are nearly as high (almost 13 percent), followed by psychoactive drugs (7 percent). In the nursing home, where patients are older and more frail than the general elderly population, the pattern has been shown to be quite different. In one study in that setting, 18 percent of prescriptions were for psychoactives, 15 percent for cathartics, and 6 percent for cardiac drugs (DPH, 1976).
Medications—Appropriateness

Table 3 and Appendix B give information on each study’s methods. Column 1 of Table 3 notes the medication class studied. Population sample size is specified by the number of subjects using the medication, the number of medications used or prescribed, or the number of pharmacies surveyed. Population source identifies the sampling site (outpatient, inpatient, or nursing home). If stated by the authors, we note whether the approach they used to judge appropriateness was benefit-risk or benefit-cost; if no statement regarding the model of appropriateness was made, we categorized the model as implicit. We also determined whether the author’s definition of inappropriate use included overuse, underuse, or misuse; if the definition was not evident, we expressed the results as “overall appropriateness.” In addition, we summarized the data on the percentage of inappropriateness of use found by each study. If actual numbers were not given, we estimated the results qualitatively. Finally, we assigned a methodologic score for each study, which was based on the scoring methods described above.

Four of the studies reported their results for “all” medications rather than for specific medication categories. Although the remaining 15 studies did contain data on specific classes of medications, they studied only three of the eight classes that constitute 50 percent of medication use in the United States. Just over 40 percent of the studies collected data on the medication use of non-institutionalized, ambulatory persons, with subjects identified through their use of medical outpatient services, by computer-based reimbursement records, by pharmacy records, or by a sample of convenience derived from those participating at senior centers. Five studies examined hospitalized populations and two examined those in nursing homes.

Eleven of the 19 studies examined overuse of medication. (In some cases the distinction between overuse and misuse was not clear, and we re-categorized those studies according to our criteria rather than the statement of the authors.) For example, studies of psychoactive drugs showed that between 7 percent and 51 percent of sedatives, hypnotics, antidepressants, and antipsychotics are overused. Antianxiety medication was the most overused. Another study found that 22 percent of emergency room antibiotic prescriptions were inappropriately overused (Bernstein et al., 1982).

In the five studies that examined underuse, inappropriateness was determined in one of two ways. In one method, subjects were identified by diagnosis, and their medical therapy was analyzed for appropriateness by reviewing the medical record (e.g., for subjects who came to an emergency room for treatment of an infectious disease and did not
<table>
<thead>
<tr>
<th>Medication Class</th>
<th>Population Size\textsuperscript{a}</th>
<th>Population Source</th>
<th>Approach to Measurement</th>
<th>Type of Inappropriateness</th>
<th>Percentage of Inappropriateness</th>
<th>Quality Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>573</td>
<td>Hospitalized</td>
<td>Implicit</td>
<td>Overall</td>
<td>5.3% (on admission)</td>
<td>4</td>
</tr>
<tr>
<td>2. 389, 981\textsuperscript{b}</td>
<td>Outpatient</td>
<td>Benefit-risk</td>
<td>Misuse/overuse</td>
<td>33% overuse; 18% misuse</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>3. 30,000\textsuperscript{b}</td>
<td>Outpatient</td>
<td>Implicit</td>
<td>Misuse</td>
<td>50% misuse</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>4. 948\textsuperscript{b}, 20\textsuperscript{c}</td>
<td>Outpatient</td>
<td>Benefit-risk</td>
<td>Overuse</td>
<td>37% overuse</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Psychoactive</td>
<td>5,902</td>
<td>Nursing home</td>
<td>Implicit</td>
<td>Misuse &amp; overuse</td>
<td>Large misuse &amp; overuse</td>
<td>5</td>
</tr>
<tr>
<td>6. 578</td>
<td>Outpatient</td>
<td>Implicit</td>
<td>Misuse &amp; overuse</td>
<td>7% overuse; 44% misuse</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>7. Large</td>
<td>Not stated</td>
<td>Implicit</td>
<td>Overuse &amp; underuse</td>
<td>Large underuse &amp; overuse</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>8. 46</td>
<td>Outpatient</td>
<td>Implicit</td>
<td>Misuse &amp; overuse</td>
<td>11% misuse; 51% overuse</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Anti-infectives</td>
<td>540</td>
<td>Nursing home</td>
<td>Implicit</td>
<td>Underuse</td>
<td>67% underuse</td>
<td>3</td>
</tr>
<tr>
<td>10. Large</td>
<td>Theoretical</td>
<td>Benefit-cost</td>
<td>Underuse</td>
<td>65%–80% underuse</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>11. 1,185</td>
<td>Emergency room</td>
<td>Implicit</td>
<td>Overuse &amp; underuse</td>
<td>22% inapprop. use, mostly overuse</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>12. 1,021</td>
<td>Hospitalized</td>
<td>Implicit</td>
<td>Misuse</td>
<td>69% misuse, prophylactic antibiotics</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>13. 267</td>
<td>Hospitalized</td>
<td>Implicit</td>
<td>Misuse</td>
<td>74% misuse, prophylactic courses</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>14. 620</td>
<td>Emergency room</td>
<td>Implicit</td>
<td>Overuse &amp; underuse</td>
<td>6% underuse; 17% overuse</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>15. 1,900</td>
<td>Outpatient</td>
<td>Implicit</td>
<td>Misuse</td>
<td>44% misuse in elderly</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>16. Not stated</td>
<td>Hospitalized</td>
<td>Benefit-cost</td>
<td>Misuse</td>
<td>60–90% misuse</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>17. 78\textsuperscript{b}</td>
<td>Hospitalized</td>
<td>Implicit</td>
<td>Misuse &amp; overuse</td>
<td>17% overuse; 37% misuse</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>GI</td>
<td>51</td>
<td>Outpatient</td>
<td>Benefit-risk</td>
<td>Overuse &amp; misuse</td>
<td>35% overuse; 31% misuse</td>
<td>4</td>
</tr>
<tr>
<td>19. 105\textsuperscript{b}</td>
<td>Outpatient</td>
<td>Implicit</td>
<td>Misuse &amp; overuse</td>
<td>25% overuse; 71% misuse</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}The population size is the number of people using medications, except where otherwise noted.

\textsuperscript{b}Number of medications.

\textsuperscript{c}Number of pharmacy sites.
receive antibiotics, the authors determined whether the failure to treat was appropriate; for those with major depressive symptoms or those with generalized anxiety, the authors determined if appropriate therapy was instituted). In another method, a population was identified as being entirely in need of treatment (e.g., failure to vaccinate for influenza was considered underuse in an elderly, nursing home population).

One study found that misuse of psychoactives occurred in 44 percent of cases, because clinicians failed to develop an adequate treatment plan (Wells et al., 1988). In two examples of misuse, cimetidine was prescribed for the wrong duration of time (Ulasek et al., 1984; Stander and Yates, 1988). Two studies examined the misuse of antibiotics when used as prophylaxis for surgery. One (Shapiro, 1979) found that antibiotic prophylaxis was given for more than two days in 93 percent of cases, although two days is the recommended duration of therapy. In a study of tetanus immunization, over half of the therapy given to persons over age 60 was inappropriate because the wrong vaccine was used (Mullooly, 1984).

As noted above, over 70 percent of the studies used implicit criteria. In 13 of the studies, the authors failed to state the criteria they used or their qualifications as judges to determine appropriateness. In 14 papers, the authors did not state the reproducibility of their determinations of appropriateness or the degree to which the validity of their assessment instruments was tested. Finally, in fewer than two-thirds of the 19 studies were subjects selected to be representative of a large community. Many of the studies were conducted at a single site or in a small localized area.

To summarize our results, we compiled the findings from the 10 best papers (those that scored 6 or above, i.e., the mean of all scores). The data from this compilation suggest that 7 to 37 percent of medication is overused in outpatients, and 44 percent to 90 percent is misused. The most that we can say about the underuse of medication is that, at least for some groups of patients with mental health symptoms, the underuse of psychoactive drugs may be substantial. None of these "best" papers studied the underuse of medication use in the nursing home.

Finally, we compared what is known about inappropriate use with what is known about expenditures. Table 4 lists, in order of expenditure, the eight classes of medication that constitute over 50 percent of the total expenditures for medication among the elderly. Comparing the small number of studies that have been done on appropriateness with the expenditures associated with these medications highlights the inadequacy of our current knowledge. What little we do know about
appropriateness suggests that as much as three-quarters of the medication prescribed is not prescribed correctly, and that we are overusing and misusing drugs. We know even less about the rate of underuse of medication.

Table 4
APPROPRIATENESS AND EXPENDITURES FOR MEDICATIONS BY MEDICATION CLASS

<table>
<thead>
<tr>
<th>Medication Class</th>
<th>Expenditure (Millions $)</th>
<th>Number of Studies</th>
<th>Range of Inappropriateness (Overuse and Misuse)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-infectives</td>
<td>1,882</td>
<td>9</td>
<td>22%-90%</td>
</tr>
<tr>
<td>Cardiac</td>
<td>1,729</td>
<td>0</td>
<td>No data</td>
</tr>
<tr>
<td>Psychosactives</td>
<td>1,011</td>
<td>4</td>
<td>7%-51%</td>
</tr>
<tr>
<td>Antiarthritis</td>
<td>805</td>
<td>0</td>
<td>No data</td>
</tr>
<tr>
<td>Analgesics</td>
<td>764</td>
<td>0</td>
<td>No data</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>556</td>
<td>2</td>
<td>33%-71%</td>
</tr>
<tr>
<td>Diuretics</td>
<td>566</td>
<td>0</td>
<td>No data</td>
</tr>
<tr>
<td>Cold/cough</td>
<td>478</td>
<td>0</td>
<td>No data</td>
</tr>
<tr>
<td>All drugs</td>
<td>25,000</td>
<td>14</td>
<td>Up to 51% overuse</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Up to 90% misuse</td>
</tr>
</tbody>
</table>
IV. DISCUSSION

This report examines the appropriateness of acute services, namely, the use of visits, hospitalizations, procedures, and medications. We deliberately excluded from this review mental health services (except in connection with medications), long-term care services (except for medications), and dental health services. Our findings do not therefore apply to these services.

When we began this work to describe what is known about the level of the appropriateness of acute care rendered to people 65 years of age and older, we knew that the literature was sparse. However, we were not adequately prepared for both the lack of information and the relatively low quality of the studies that we found. For the most part, information was either out of date or not generalizable. Thus, the first question that faces policymakers, patients, and physicians is whether one wants valid and generalizable information about the appropriateness of the acute services rendered in the U.S. health system to the elderly. If so, it will require the annual investment of tens of millions of dollars. These funds would be required to establish appropriateness guidelines for 50 to 100 procedures, to update them no less frequently than every three years, and to apply them to a sample of the U.S. population on a regular basis so that sound, epidemiologic conclusions on the appropriateness of care in the United States by region, gender, age, and race can be made. From the perspective of researchers, who are admittedly biased, such an investment seems warranted. Planning for a better health care system cannot occur with the almost complete absence of information about the appropriateness of care.

From a research perspective, both the lack of and low quality of the information contained in the literature limited the power of the analytical techniques that we established at the beginning of our research endeavor to yield much information. In particular, we had hoped to perform an indepth meta-analysis of the literature.

We first specified a sampling frame to identify relevant literature, and then developed a rating scale by which to judge each article’s methodologic adequacy. We had hoped to identify quickly those articles that met minimally acceptable methods criteria, and to spend most of our time commenting about the relationship between the studies’ methods and results. For example, we were interested in determining if valid and reliable methods for measuring appropriateness label more patients as having received inappropriate care than if the methods were
invalid or unreliable. Does the level of appropriateness of care vary by whether the information for judging appropriateness is obtained from the patient, the medical record, or the physician? If so, which is more valid? How important are patients' utilities in judging appropriateness? Do different approaches (e.g., benefit-risk, benefit-cost) for determining appropriateness produce different levels of inappropriate care? If explicit criteria are used (compared with implicit criteria), does the conclusion about appropriateness of care change? If data are collected prospectively (as opposed to retrospectively), do appropriateness results change? These are only some of the research questions that could and should be addressed as part of a major effort to identify the level of appropriateness of care.

What, then, did we find? Even given the state of the literature, it is possible to comment, for at least a few procedures, about the level of inappropriate care delivered in the United States. Almost every study found at least double-digit levels of inappropriate care. Those studies that looked at overuse in the hospital, of procedures, or of medications, found evidence of such overuse in amounts that are too large to be ignored. In particular, perhaps as much as one-fifth to one-quarter of acute care services were felt to be used for equivocal or inappropriate reasons. (This statement needs to be tempered by the realization that the topics selected may not be representative of all procedures.)

The few studies that examined underuse or misuse of services also documented the existence of inappropriate care. This was especially true for the ambulatory care of chronic physical and mental conditions and mostly concerned the use of low-cost technologies (visits, some medications). Thus, we conclude that there appears to be a substantial problem in the matching of acute care services to the needs of elderly patients. This mismatch occurs both in terms of overuse and underuse, at least for the areas where research has been conducted.

This conclusion can be viewed pessimistically as an indictment of the current U.S. health care system or optimistically as a window of opportunity. Even while awaiting more data on appropriateness, much work needs to be done to decrease the extent of inappropriateness in the U.S. health care system. Depending on one's political and philosophical beliefs, that work could take the form at the policy level of increased regulation or competition, or at the clinical level of better education or dissemination of information to patients, physicians, and other health care professionals. Accomplishing any part of this agenda will be painful. On the other hand, by recognizing that a large percentage of what we currently do is not needed, it is possible to identify scenarios that would free up resources that could be used to provide better acute care as well as improved care in two areas we did not
address—long-term care and mental health care. By improving acute care, such a strategy might also reduce the prevalence of iatrogenic illness. Of course, such an agenda would take time to complete, and would be difficult to accomplish, because it would involve the translocation and retraining of many people (e.g., physicians educated in acute care might need to provide long-term care). We believe, however, that it is time to make a commitment to measuring appropriateness and using the appropriateness results to change how health resources are spent in maintaining the well-being of the elderly. In addition, experiments at a clinical or patient-doctor level that are designed to improve appropriateness of care can begin now.
Appendix A

APPROPRIATENESS OF USE OF PROCEDURES, HOSPITALS, AND AMBULATORY CARE: AN ANNOTATED BIBLIOGRAPHY

PROCEDURES

1. Winslow et al. (1988a).

Health service: Coronary artery bypass surgery (CABS).

Approach: Benefit-risk with appropriateness standards based on expert consensus ratings of 488 indications and literature review.

Methods: Examined a random sample of admissions in which CABS was performed from three randomly chosen hospitals in a western state. Conducted medical record review to determine indication for the procedure. Study patients were assigned an indication for the procedure. Indications were graded on a 1 to 9 scale of appropriateness determined by a panel of expert physicians.

Population: Random sample of all Medicare patients who received CABS in three randomly chosen hospitals in a western state.


Conclusions: Fourteen percent of the surgeries were performed for inappropriate reasons, 30 percent for equivocal (neither appropriate or inappropriate) reasons, and appropriateness varied by hospital from 37 percent to 78 percent. The percentages of appropriate and inappropriate cases for patients younger or older than 65 years were similar. The mean appropriateness score showed slightly more appropriate use in the elderly.

Overall quality assessment of article: 14/14 = above average.¹

¹Ranking of each study was based on the mean score within each category of care (“average” was two points below to two points above the mean score).

Health service: Coronary angiography.

Approach: Benefit-risk with appropriateness standards based on expert consensus ratings of 300 indications and literature review.

Methods: Examined Medicare physician claims data to determine population-based rates of use for coronary angiography in 13 sites, then selected three sites with the use rates at both extremes and conducted medical record review to determine indication for the procedure. Indications were graded on a 1 to 9 scale of appropriateness determined by a panel of expert physicians.

Population: Medicare patients who received coronary angiography in high-use and low-use geographic regions (three sites).

Time period: 1981.

Conclusions: Variations in use were not due solely to appropriateness of the procedures. There was a high level of inappropriate use of coronary angiography (17 percent), and an additional 9 percent were equivocal. Patients in the high-use site were older and less intensively medically treated than patients in either of the low-use sites. The most common subgroup of inappropriate cases were patients without angina who had not undergone exercise testing.

Overall quality assessment of article: 13/14 = above average.


Health service: Carotid endarterectomy.

Approach: Benefit-risk with appropriateness standards based on expert consensus ratings of 864 indications and literature review.

Methods: Examined Medicare physician claims data to determine population-based rates of use for carotid endarterectomy in 13 sites, then selected three sites, with two sites representing rates at both extremes, and conducted medical record review to determine indication for the procedure. Indications were graded on a 1 to 9 scale of appropriateness determined by a panel of expert physicians.
Population: Medicare patients who received carotid endarterectomy in high-use, medium use, and low-use geographic regions (three sites).

Time period: 1981.

Conclusions: Carotid endarterectomy was overused (32 percent inappropriate, 32 percent equivocal), and because the average complication rate for this procedure is high, use of it even for appropriate indications might not be justified because the risks would probably outweigh the benefits.

Overall quality assessment of article: 13/14 = above average.


Health service: Coronary angiography, carotid endarterectomy, and upper gastrointestinal tract endoscopy. (Because results of angiography and carotid endarterectomy are reported elsewhere [Chassin 1987a and Winslow 1988b], this abstract concerns UGI endoscopy.)

Approach: Benefit-risk with appropriateness standards based on expert consensus ratings of 1,058 indications and literature review.

Methods: Examined Medicare physician claims data to determine population-based rates of use for endoscopy in 13 sites, then selected three sites with the rates at both extremes and conducted medical record review to determine the appropriateness of the indications for the procedure. Indications were graded on a 1 to 9 scale of appropriateness determined by a panel of expert physicians.

Population: Medicare patients who received endoscopy in high-use and low-use geographic regions (three sites).

Time period: 1981.

Conclusions: Variations in use were not due to appropriateness. There was a high level of inappropriateness for endoscopy; 17 percent of the procedures were found to be inappropriate and 11 percent were equivocal. (These results also appear in Kahn et al., 1988.)

Overall quality assessment of article: 13/14 = above average.
5. Merrick et al. (1986).

**Health service:** Carotid endarterectomy.

**Approach:** Benefit-risk with appropriateness standards based on expert consensus ratings of 864 indications and literature review.

**Methods:** Examined the medical records of all patients undergoing carotid endarterectomy in five California Veterans Administration (VA) Medical Centers. Indications were graded on a 1 to 9 scale of appropriateness determined by a panel of expert physicians.

**Population:** VA patients of all ages who received carotid endarterectomy.

**Time period:** 1981.

**Conclusions:** Carotid endarterectomy was clearly inappropriate in 13 percent of all patients, equivocal in 32 percent, and appropriate in 55 percent.

**Overall quality assessment of article:** 9/14 = average.


**Health service:** Appendectomy, cataract surgery, cholecystectomy, coronary angiography, uterine dilation and curettage, surgical repair of hiatal hernia, abdominal hysterectomy, vaginal hysterectomy, lumbar disk excision, meniscectomy, tonsillectomy, adenoidectomy, and adenoidectomy.

**Approach:** Mostly implicit.

**Methods:** Prospective medical record review by nurses who used an explicit set of general indications to determine if any indication for doing the procedure was met. Any case that failed to meet at least one indication was referred to a physician who could, using implicit judgment, override the inappropriateness judgment. Case could also be discussed with the admitting physician.

**Population:** 4,850 Medicare and Medicaid patients admitted to all Colorado hospitals during a six-month study period.
Time period: June 1, 1978, to December 1, 1978.

Conclusions: Ninety-seven percent of the cases met at least one indication for each procedure or were justified by physician review.

Overall quality assessment of article: 6/14 = average.

7. Melton et al. (1982).

Health service: Hip arthroplasty.

Approach: Implicit and the rates of use were assumed to be optimal.

Methods: All hip arthroplasties done in a single county served by the Mayo Clinic were identified and medical records were reviewed to determine age- and sex-adjusted incidence. Appropriateness was not assessed.

Population: 381 operations performed on 321 patients in Olmsted County, Minnesota.


Conclusions: If all hip arthroplasties are assumed to be performed for appropriate reasons in Olmsted County, then this procedure might be underused in the United States because national rates are much less than are the rates found in Olmsted County.

Overall quality assessment of article: 4/14 = below average.


Health service: Preoperative laboratory screening.

Approach: Implicit with explicit criteria defined by the authors.

Methods: Criteria both for ordering lab tests and for abnormal lab values were set up before the study. Medical records were reviewed by physicians for inappropriately done tests that had abnormal results.

Population: A random sample of 6,200 preoperative tests done on 2,000 patients admitted during the study period.

Conclusions: Sixty percent of tests were inappropriate; of those tests that were done inappropriately, 0.22 percent revealed abnormalities that might have influenced perioperative management, and chart review of these cases revealed that the abnormal test results did not influence perioperative management and did not adversely affect patient outcomes.

Overall quality assessment of article: 3/14 = below average.


Health service: Cardiac pacemaker implantation.

Approach: Implicit with implicit criteria defined by the authors.

Methods: Retrospective medical record review by physicians and nurses.

Population: 382 Medicare implantations in 30 hospitals in Philadelphia County.

Time period: January 1 to June 30, 1983.

Conclusions: Authors found that 44 percent of pacemaker implantations were clearly appropriate, 20 percent were clearly inappropriate, and 36 percent were possibly appropriate.

Overall quality assessment of article: 3/14 = below average.


Health service: Cardiac pacemaker implantation.

Approach: Implicit.

Methods: Reviewed several hospital files recording pacemaker implantation as well as previous research studies.

Population: varies.
Time period: varies.

Conclusions: Between 30 percent and 75 percent of pacemaker implantations are inappropriate.

Overall quality assessment of article: 2/14 = below average.

HOSPITAL SETTING

1. Siu et al. (1986).

Health service: Hospital admission and hospital day.

Approach: Benefit-risk with explicit criteria based on the Appropriateness Evaluation Protocol or AEP.

Methods: Medical record review in a randomized trial of health insurance plans.

Population: 1,132 hospitalized adults (age 18 to 61) representative of the civilian, noninstitutionalized population from six sites representative of four census regions. This study was included because some subjects were 50 years of age or older.


Conclusions: Health insurance plan (free versus cost-sharing) did not influence the appropriateness of hospital admission or hospital days. The percentage of inappropriate admissions varied among six geographic sites (10 percent to 35 percent), but the areas with low admission rates did not have low proportions of inappropriate admissions. Thirty-five percent of hospital days were avoidable. A substantial fraction of hospitalization was potentially avoidable and cost-sharing did not selectively reduce inappropriate hospitalization.

Overall quality assessment of article: 12/14 = above average.

**Health service:** Hospital days of care and admissions.

**Approach:** Implicit with explicit criteria (AEP).

**Methods:** Medical record review by trained nurses and physicians of randomly selected patients admitted on randomly selected days.

**Population:** Medicare and Medicaid patients in 25 hospitals in four geographic regions.

**Time period:** Representative sample of each of the days of the week within each of the four seasons of 1980.

**Conclusions:** Twelve percent to 28 percent, depending on geographic region, of all hospital admissions, and 21 percent to 37 percent of days of care were found to be inappropriate. For all of the regions, one in five admissions and one in four days of care were felt to be inappropriate. For inappropriate hospital admissions, hospital type (teaching versus large nonteaching versus small) was found to be significant; however, there were no consistent differences among all three hospital types across all four regions. Overall, 43 percent of the inappropriate admissions resulted from an inappropriate site of care (e.g., hospital admissions for diagnostic or therapeutic services that could have been rendered at an ambulatory site). For inappropriate days of care, all four regions were more likely to have inappropriate days of care at the end of the length of stay. Ninety-two percent of inappropriate days occurred because reviewers felt that the site of care was inappropriate (e.g., patients could have received care on an ambulatory or nonacute institutional basis). The majority of inappropriate use was due to hospital and physician related reasons and not to patient factors. Variations in inappropriate use appeared to correspond to the variation in total hospital use in the four regions.

**Overall quality assessment of article:** 6/14 = average.


**Health service:** Hospital day of care.

**Approach:** Implicit with explicit criteria.
Methods: Intervention study to increase appropriateness. Time series design. Appropriateness judgment based on medical record review by trained nurses. Intervention was the concurrent feedback of individual physician's appropriateness ratings compared to peer norms.

Population: All 6,574 Medicare patients admitted to six hospitals within four adjacent counties of Maryland during the study time period.

Time period: 1983.

Conclusions: Seven percent of admissions failed (either inappropriate admission or day of stay) during baseline period; rate decreased over 50 percent during the intervention period.

Overall quality assessment of article: 8/14 = average.


Health service: Hospital day of care.

Approach: Explicit, examining: (1) appropriateness of the hospital setting using explicit criteria based on the AEP; and (2) Medical appropriateness using implicit criteria based on a 14 member physician panel which used a 9 point scale of appropriateness.

Methods: Medical record review of hospitalized patients randomized into a fee-for-service (FFS) or health maintenance organization (HMO).

Population: A random sample of families representative of the civilian, noninstitutionalized population under 62 years of age in the Seattle, Washington, area.


Conclusions: Thirty-two percent of fee-for-service admissions were discretionary (i.e., did not meet AEP criteria or met criteria but therapy that justified hospitalization was not medically necessary), as well as 27 percent of HMO admissions.

Overall quality assessment of article: 8/14 = average.

**Health service:** Hospital days of care.

**Approach:** Implicit with explicit criteria using Medicare level-of-care criteria.

**Methods:** Intervention study to increase appropriateness. Experimental design with patients randomly assigned to one of four intervention groups. Appropriateness judgment based on medical record review by trained nurses. Interventions were four feedback strategies including: direct feedback to the attending physician from the nurse reviewer, indirect feedback, no feedback, and judgmental feedback.

**Population:** Medicare patients admitted to four hospitals in San Francisco.

**Time period:** Two months in 1981.

**Conclusions:** Before intervention, 6 percent of hospital days felt to be inappropriate. Judgmental and direct feedback decreased inappropriate days of care when compared to indirect or no feedback.

**Overall quality assessment of article:** 3/14 = below average.


**Health service:** Hospital care for breast cancer patients.

**Approach:** Benefit-risk with explicit “criteria map” based on standard texts and expert review. Determined the appropriateness of breast cancer management (or lack of) including diagnostic evaluation, staging of the tumor, and therapy.

**Methods:** Medical record review of a random sample of elderly and nonelderly patients stratified on age, comorbidity, hospital, and physician.

**Population:** 374 patients 50 years and older with a histologically confirmed adenocarcinoma of the breast were sampled from seven Southern California hospitals selected to vary in patient socioeconomic status, volume of cancer patients, and university affiliation.

Conclusions: There was both underutilization of appropriate treatment as well as overutilization of inappropriate treatment; 17 percent of women 70 and older did not receive appropriate surgery and 4 percent of those age 50 to 69 did not.

Overall quality assessment of article: 11/14 = above average.

AMBULATORY CARE


Health service: Ambulatory care for five common conditions.

Approach: Implicit with criteria for “standards of minimally acceptable care.” (Criteria were outcome [e.g., blood pressure, <105 mm. Hg] or process [e.g., flu shot given or audiometry performed].)

Methods: Cases were identified from the Health and Nutrition Examination Survey (HANES) using objective criteria for five conditions: hypertension, angina, hearing impairment, shortness of breath on exertion, and depression.

Population: An elderly (age 65 to 74) subset of this population-based sample totaling 1,226 persons.


Conclusions: Inappropriate care ranged from 26 percent to 80 percent depending on the condition. Hypertension care was the most appropriate, depression care the least. Inappropriate care was primarily physician-related (e.g., the physician did not evaluate or treat the patient’s problem appropriately) rather than patient-related (e.g., the patient with the tracer symptoms did not seek a physician visit). Poverty was the strongest predictor of inappropriate care for all of the tracer conditions except depression.

Overall quality assessment of article: 7/14 = average.
Appendix B

APPROPRIATENESS OF USE OF MEDICATIONS: AN ANNOTATED BIBLIOGRAPHY

MEDICATION USE

Drug: All medications.

Approach: Implicit.

Method: The review process was not stated in detail. Case records were examined on admission, daily during hospitalization, and on discharge. Prescriptions were analyzed for (1) well-documented history of adverse reaction, (2) contraindication due to diagnosis, and (3) potentially harmful drug-drug interactions of clinical significance.

Population: 573 elderly patients admitted to hospital.

Time period: January 1983.

Conclusions: Patients received 2.1 drugs on admission, 5.5 during hospital stay, 3.5 on discharge. Sixty-five errors (affecting 59 patients) were noted in the 1,226 admission prescriptions (5.3 percent); 135 errors (affecting 120 patients) were seen in the 4,934 hospital prescriptions (27 percent).

Overall quality assessment of article: 4/14 = average.


Drug: All medications.

Approach: Benefit-risk.

Method: The authors screened for medication use. They categorized medications as: (1) of no therapeutic value, (2) of demonstrated therapeutic value, or (3) fixed-dose combinations of both groups.
**Population:** 2 sites—981 prescriptions from an outpatient clinic of the Spanish Social Security System, and 389 randomly chosen individuals from a pensioners' club.

**Time period:** One month, year not stated (outpatient clinic); 5 weeks in 1978 (pensioners' club).

**Conclusions:**

**Overall quality assessment of article:** 4/14 = average.

3. Helling et al. (1982).

**Drug:** All medications.

**Approach:** Implicit.

**Method:** The authors surveyed 100 community pharmacies in Iowa; they reviewed 30,000 prescriptions reimbursed under Medicaid. They used previously published criteria for review (Knapp). Their criteria for inappropriateness were use of irrational drug combinations, prescription of incorrect dose, or number of pills ordered.

**Population:** Noninstitutionalized, 30,000 prescriptions.

**Time period:** 1976.

<table>
<thead>
<tr>
<th>Medications Used by “Value” and Age</th>
<th>≥ 65</th>
<th>&lt; 65</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drugs of no value</td>
<td>162 (33 %) (^a)</td>
<td>80 (19 %)</td>
</tr>
<tr>
<td>Demonstrated value</td>
<td>239 (49 %)</td>
<td>263 (61 %)</td>
</tr>
<tr>
<td>Fixed-dose</td>
<td>89 (18 %)</td>
<td>86 (20 %)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>490</strong></td>
<td><strong>429</strong></td>
</tr>
</tbody>
</table>

**NOTE:** Significantly, more drugs of "no value" were used by the elderly.

\(^a\)p < 0.001, comparison between age groups.
Conclusions: Seventeen percent of prescriptions were for irrational drug combinations, 31 percent were prescribed in inappropriate quantity, 9 percent were for inappropriate dose; 50 percent of prescriptions violated at least one of their criteria.

Overall quality assessment of article: 8/14 = average.


Drug: All medications.

Approach: Benefit-risk.

Method: Effectiveness ratings were developed by five clinicians in a division of clinical pharmacy. The authors applied these criteria to sales data collected over a period of three years. They surveyed 20 pharmacies plus a large drug wholesaler; 948 products were assessed.

Population: Spain, no further data.


Conclusions:

The percentages of medications of doubtful, no, or unacceptable effectiveness were 58 percent, 54 percent, and 47 percent for the three time periods. Only slight improvement over the decade was seen.

Overall quality assessment of article: 9/14 = above average.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>High effectiveness</td>
<td>32%</td>
<td>36%</td>
<td>41%</td>
</tr>
<tr>
<td>Relative effectiveness</td>
<td>10%</td>
<td>11%</td>
<td>12%</td>
</tr>
<tr>
<td>Doubtful effectiveness</td>
<td>2%</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>No effectiveness</td>
<td>19%</td>
<td>18%</td>
<td>23%</td>
</tr>
<tr>
<td>Unacceptable effectiveness</td>
<td>37%</td>
<td>33%</td>
<td>21%</td>
</tr>
</tbody>
</table>
5. Ray et al. (1980).

**Drug:** Antipsychotic medications.

**Approach:** Implicit.

**Method:** Authors reviewed the medical literature for standards of appropriate use. They reviewed medical records of nursing home subjects. They also collected data on nursing homes and physician characteristics for use in a regression analysis.

**Population:** 5,902 Medicaid nursing home residents.

**Time period:** July 1975 to June 1976.

**Conclusions:** Forty-three percent of all residents received antipsychotics. Use correlated positively with size of nursing home and size of physician’s nursing home practice. Authors suggested large overuse of these medications.

**Overall quality assessment of article:** 5/14 = average.


**Drug:** Psychoactives.

**Approach:** Implicit.

**Method:** Authors developed appropriateness criteria, which were pre-tested. New patients were chosen and their records reviewed for medication use, indications, and treatment plans.

**Population:** Fifteen sites, 578 new patients, in outpatient teaching hospital general medical group practices.

**Time period:** September to December 1981.

**Conclusions:** Fifty-six percent had adequate treatment plans documented for sedative/hypnotic use; 11 percent taking minor tranquilizers had adequate plans. More than 30 percent of those on antidepressants did not have adequate follow-up, and 14 percent received more than one minor tranquilizer at the same time. Ninety-three percent of
those using minor tranquilizers had an appropriate indication for use. Summary: misuse of these medications was high.

**Overall quality assessment of article:** 11/14 = above average.


**Drug:** Psychoactives.

**Model:** Implicit.

**Method:** Authors used pretested criteria of mental distress. Subjects were interviewed for symptoms and their medical treatment was reviewed.

**Population:** Nationwide survey.

**Time period:** 1979.

**Conclusions:** Patients with major depression received antidepressants 11 percent of the time and antianxiety medications 23 percent of the time; those with generalized anxiety were treated 19 percent of the time with antianxiety drugs and 3 percent of the time with antidepressants. "The majority of persons with serious psychiatric disorders still do not receive treatment or the most appropriate treatment."

**Overall quality assessment of article:** 6/14 = average.


**Drug:** Psychoactives, antidepressants.

**Approach:** Implicit.

**Method:** Authors enrolled a randomized sample of noninstitutionalized persons over age 70 in New Zealand in a longitudinal study and recorded medication use. Six percent (45) were found to be taking antidepressants. They also administered a health status and psychological well-being assessment. Those taking tricyclics were matched with subjects who were not.

**Population:** Forty-five subjects taking antidepressants, all receiving health care at a single health center.
Time period: Not stated.

Conclusions: Of the 45 taking antidepressants, 20 (44 percent) had evidence of a major depressive event justifying treatment. Five of these 20 were continued on antidepressants without evidence that there was continued need (misuse by our definition), and 23 (51 percent) had antidepressants started without evidence of a depressive episode requiring treatment. (In two cases a determination could not be made.)

Overall quality assessment of article: 5/14 = average.


Drug: Influenza vaccine.

Approach: Implicit.

Method: The authors used previously published standards for influenza vaccination. They reviewed vaccination histories of nursing home patients and sent questionnaires to nurses and physicians regarding attitudes toward flu vaccination.


Time period: 1983.

Conclusions: Underuse of flu vaccine was widespread, with 33 percent of those who should have been vaccinated being vaccinated. A new policy of vaccination was able to achieve 90 percent vaccination rates.

Overall quality assessment of article: 3/14 = below average.


Drug: Pneumococcal vaccine.

Approach: Benefit-cost.

Method: Authors used health adjusted life years in a conceptual model with assumptions based on literature for risk, costs, duration of immunity, etc. They also reviewed vaccine sales to determine current use.
Population:  Broad population, conceptual.

Time period:  1983.

Conclusions:  The use of pneumococcal vaccine is cost effective if the duration of immunity is eight years. The authors showed that vaccine coverage ranged from 20 percent to 35 percent of target groups, depending on the definition of high-risk groups. Thus, 65 percent to 80 percent of target groups were undertreated. They state that “current levels of vaccination appear too low considering the potential health benefits and cost-effectiveness.”

Overall quality assessment of article:  10/14 = above average.


Drug:  Antibiotics.

Approach:  Implicit.

Method:  Two of the authors reviewed medical records. Antibiotic therapy was judged as: not necessary, wrong drug, wrong dose, wrong route, or no drug given when therapy was indicated. No indication of whether disagreement between the two assessors occurred or how it was resolved.

Population:  1,185 emergency room visits; single site (university teaching hospital).

Time period:  July 1978 to December 1978, 20 randomly chosen days.

Conclusions:  Seventy-eight percent of antibiotic prescriptions were considered appropriate, and 80 percent of inappropriate use was due to overuse. In 18 percent of the inappropriate use, the wrong antibiotic was given. Forty-nine percent of the regimens considered unnecessary involved a dermatologic diagnosis, and more than half were for lacerations or abrasions. Use of antibiotics was thought to be inappropriate in 40 percent of cases in which skin or soft tissues were involved. Moreover, in 72 percent of cases in which an antibiotic was used in the eye, an antibiotic was considered unnecessary or the wrong agent.

Overall quality assessment of article:  3/14 = below average.

**Drug:** Antibiotics.

**Model:** Implicit.

**Method:** Authors established two explicit criteria for surgery patients: (1) antibiotic prophylaxis should be given within 48 hours before skin incision; (2) prescription should be limited to < 72 hours after surgery. They then surveyed use of antibiotic prophylaxis in surgical patients.

**Population:** Approximately 50 at each of 20 sites (total = 1,021).

**Time:** Started July 1979.

**Conclusion:** Only the data before the authors’ educational intervention are presented. They found that 41 percent of antibiotics were given before 48 hours before surgery; the average length of prescription was 3.4 days; 31 percent of prophylactic courses were appropriate by both criteria.

**Overall quality assessment of article:** 8/14 = average.


**Drug:** Antibiotics.

**Model:** Implicit.

**Method:** Authors established duration of treatment more than 48 hours after surgery as the sole criterion for defining use as inappropriate. They reviewed the patients’ hospital records, carefully establishing that antibiotic use was for prophylaxis only.

**Population:** 267 patients from 20 hospitals in Pennsylvania.

**Time:** Ten days in 1974.

**Conclusions:** Prophylactic antibiotic use represented 30 percent of all antibiotic use in these hospitals. Ninety-three percent of all patient days of prophylaxis represented administration for more than two days; 74 percent of all prophylactic treatment courses were for more than
two days, and 49 percent were for more than four days, twice the recommended duration of therapy.

**Overall quality assessment of article:** 8/14 = average.


**Drug:** Antiinfective (tetanus prophylaxis).

**Model:** Implicit.

**Method:** Authors examined the adequacy of antitetanus prophylaxis given to people presenting to emergency rooms with soft-tissue injuries. They used previously published recommendations for therapy as their criteria and validated data collection instruments to assess the therapy.

**Population:** 620 visits to six emergency rooms.

**Time:** July 1980 to September 1982.

**Conclusion:** Twenty-three percent of patients were treated incorrectly: 6 percent were undertreated and 17 percent were overtreated.

**Overall quality assessment of article:** 9/14 = above average.


**Drug:** Antiinfective (tetanus immunization).

**Model:** Implicit.

**Method:** Authors screened 1,900 randomly selected charts of HMO members for tetanus immunization. They found that 39 percent had received at least one injection during 10 years of eligibility. They used published criteria to establish the type of immunization that should have been given and then determined if the correct vaccine was given.

**Population:** 1,900 randomly selected charts from one HMO for those continuously eligible for 10 years.

**Time:** Patients enrolled in HMO from October 1969 through September 1979.
**Conclusion:** Forty-four percent of persons over age 60 and 29 percent of those under age 60 received only tetanus toxoid rather than the combined tetanus-diphtheria toxoids, as recommended by the Centers for Disease Control.

**Overall quality assessment of article:** 7/14 = average.


**Drug:** Antibiotics (clindamycin, cefazolin, metronidazole).

**Model:** Benefit-cost.

**Method:** Article presents baseline data from a study that examined the effectiveness of an educational order form on improving appropriate prescribing of three antibiotics. Authors reviewed all prescriptions for the three target antibiotics in one Boston hospital to determine the dosing schedules ordered. The criteria were generated from literature review and expert opinion.

**Population:** Exact number of patients not stated in article.

**Time:** Not stated.

**Conclusions:** Misuse due to too-frequent dosing accounted for 90 percent, 60 percent, and 75 percent of patient days for therapy with clindamycin, cefazolin, and metronidazole, respectively.

**Overall quality assessment of article:** 7/14 = average.


**Drug:** Antibiotic (cefoxitin).

**Model:** Implicit.

**Method:** Authors retrospectively reviewed cefoxitin use through chart review at one Canadian teaching hospital. They presented the explicit criteria for appropriate use but did not state the source of those criteria or the authors' qualifications in generating them.

**Population:** Seventy-eight orders for cefoxitin were reviewed.
**Time:** December and January 1980; December 1980 and January 1981.

**Conclusions:** Cefoxitin was inappropriately used in 43 percent of patients: 7 percent, antibiotic therapy not necessary; 16 percent, alternative agents should have been used; 10 percent, cefoxitin was redundant to another antibiotic prescribed; 10 percent, errors in dose, scheduling, or duration of therapy.

**Overall quality assessment of article:** 3/14 = below average.


**Drug:** Cimetidine.

**Approach:** Benefit-risk.

**Method:** Authors used literature review and FDA-established criteria of appropriate use. They reviewed medical records of subjects who were prescribed cimetidine.

**Population:** Fifty-one noninstitutionalized patients from a single hospital outpatient pharmacy.

**Time period:** March 1982.

**Conclusions:** Long-term cimetidine prescription was inappropriate in 34 of 51 patients (67 percent). In 16, full-dose therapy was excessive (31 percent) (misuse by our definition). In 18 (35 percent), there was no acceptable indication for cimetidine therapy (overuse by our definition).

**Overall quality assessment of article:** 4/14 = average.


**Drug:** Gastro-intestinal (H2 blocker).

**Model:** Implicit.

**Method:** Authors reviewed the ambulatory records all patients treated at one Department of Health Services who received H2 blockers. A review panel appointed by the Quality Assurance Committee
performed the review; the credentials of the panel were not stated. Broad criteria were set to determine appropriateness.

**Population:** 103 new or renewed prescriptions written during a two-week period were reviewed.

**Time:** Two-week period in November 1986.

**Conclusions:** Twenty-five percent of those receiving medications had no appropriate indication for therapy; 71 percent were being treated with an inappropriate dosage of medication or for an inappropriate duration.

**Overall quality assessment of article:** 4/14 = average.
BIBLIOGRAPHY


Aging America. Trends and Projections, prepared by the U.S. Senate Special Committee on Aging, in conjunction with the American Association of Retired Persons, 1984.

Aging America. Trends and Projections, prepared by the U.S. Senate Special Committee on Aging, in conjunction with the American Association of Retired Persons, 1988.


