A RAND NOTE

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 a grant from the U.S. Department of Health and Human Services.

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

The RAND Publication Series: The Report is the principal publication documenting and transmitting RAND's major research findings and final research results. The RAND Note reports other outputs of sponsored research for general distribution. Publications of RAND do not necessarily reflect the opinions or policies of the sponsors of RAND research.
A RAND NOTE

N-3371-HHS

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

Supported by the U.S. Department of Health and Human Services
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 and Andrea Steiner

Departments of Medicine and Public Health UCLA Center for the Health Services, Los Angeles, and the RAND Corporation, Santa Monica, CA, U.S.A.

Accepted 20 February 1990

Summary

Over the past 30 years, an explosion in health care expenditures has occurred. Prior to 1960, health care accounted for 4.4% of the U.S. Gross National Product; today it is 11%. Before rational solutions to controlling this rise can be proposed, we must determine whether the care that we are currently paying for is appropriate to the needs of the elderly. This paper analyzes the literature regarding appropriateness of acute care provided to the elderly.

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, and 19 articles that presented data on the appropriateness of medication use in the elderly. Virtually every study included in this review found at least double-digit levels of inappropriate care. Perhaps as much as one-fifth to one-quarter of acute 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 the existence of these phenomena. This was especially true for the ambulatory care of chronic physical and mental conditions and concerned the use of low-cost technologies (visits, preventive services, some medications). Thus, we conclude that there appears to be a substantial problem in the matching of acute services to the needs of elderly patients. This mismatch occurs both in terms of overuse and underuse, at least for areas where research has been conducted.

Address for correspondence: Professor Robert H. Brook, M.D., The RAND Corporation, 1700 Main Street, PO Box 2138, Santa Monica CA 90406-2138, U.S.A.

Introduction

Over the past 30 years a virtual explosion in health care expenditures has occurred. Before 1960, health care accounted for 4.4% of the Gross National Product (GNP) [1]. 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%) [2].

Part, although by no means all, of the increase in Medicare outlays can be attributed to demographic changes. Between 1970 and 1984 the percent of the total population 65 and older increased from 9.8 to 11.8%. At the same time, the 85 and over group has become the fastest growing segment of the elderly population.

These trends are expected to continue, and the nation's total health care bill was projected to be $640 billion dollars by 1990. Of this, $254 billion was expected to be spent on hospitals, $132 billion on physicians' bills, and $44 billion on drugs. Total Medicare costs to the federal government were projected to be $119 billion dollars [3].

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 [4,5]. Some maintain that the consequence of containing costs in the United States would be rationing such as that found in the British system [6]. Others have suggested ways to limit our health care spending by discontinuing reimbursement for many forms of acute care for the very old [7].

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, and/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 [8-11].

We suggest that, before we even begin to discuss the rationing of care according to one or more of the economic solutions noted above, 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?

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 when and which services Medicare and private insurers should cover. We would know when and how to consider patient preferences. 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).

The purpose of this paper is to review the literature that examines the appropriateness of acute care provided to the elderly and, based upon this literature review, to suggest options at both the policy and clinical levels that could promote more appropriate care. We examine the appropriateness of three kinds of care used by the elderly: hospital, ambulatory, and pharmaceutical. In performing this literature review, we were often placed in the uncomfortable position of reviewing our own work. Our intent, however, was not to determine whose work is the 'best'. Rather, using criteria developed both during our previous work in this area and during the literature review, our aim was to state what we know about the level of appropriate care.

Methods

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.

Based on a literature review, we have delineated three approaches by which to judge whether a service is appropriate. 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. The third approach relies entirely on implicit judgements, without specifying what the appropriateness rater should take into account or what he or she should disregard.

**Benefit-risk approach**  According to this approach, a service is judged to be appropriate only if the health benefits to the patient exceed the negative health consequences to the patient by a wide enough margin so 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 to be appropriate, no matter how expensive it is to provide that service [12–15].
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, or his insurance carrier. 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.

Implicit approach In the implicit approach, standards for judging appropriateness may have been used, but no clear, explicit definitions are specified. In general, the approach relies upon individual physician judgement, and it is unclear what appropriateness model the physician has in mind when he or she makes such a judgement.

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 she attaches to them. For example, implicit judgements 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.

Each of these approaches to measuring appropriateness can be used to classify a service as being underused, misused, or overused [16]. 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.

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.

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 (e.g., use of Tricyclic anti-depressants to treat insomnia [17]).
Literature review

To determine the level of appropriateness of acute care that was provided to the elderly, we conducted a MEDLINE search covering the years 1980 to 1988. Our initial search identified over 300 possibly relevant articles. We identified others by talking to experts, examining reference lists, and by scanning the tables of contents of health services research journals. This paper 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 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 users of the specific drug even if the population studied did not include those over 50.

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 ten 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.

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 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 [18]. Tests of reliability can be made by comparing results from alternate 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).

Validity refers to the extent to which a measure assesses what it purports to measure. Internal validity addresses those factors that guarantee the study’s results are ‘true’ (i.e., whether a study’s results are due to the planned intervention or to external factors). External validity refers to those aspects of a research design that
Table 1
Scoring system for rating articles on appropriateness

A. Characteristics of the approach used to measure appropriateness

1. Implicit vs. explicit approach
   0 = implicit approach
   1 = benefit-risk or benefit-cost approach

2. Type of standards used to judge appropriateness
   0 = implicit criteria: criteria not explained (e.g., based only on author’s subjective opinion)
   1 = explicit criteria used

B. Analysis by age

3. Addressed separately or as part of larger group
   0 = sample includes people over age 50, but results are not presented by age category
   1 = analysis presented by age category, 65 and over
   2 = analysis presented by subcategories of age 65 and older (e.g., 65-74, 75-84, etc.)

C. Reliability*

4. Reliability of data collection
   0 = not tested
   1 = tested

5. Reliability of judges
   0 = not tested
   1 = tested

6. Level of reliability met
   0 = not reliable or results not given
   1 = if researchers conducted reliability tests (as described in the text), they reported reliability for either data collection or judges’ consensus panels
   2 = if researchers conducted reliability tests (as described in the text), they reported reliability for both data collection and judges’ consensus panels

D. Internal validity*

7. Judgment processes
   0 = if explicit, criteria lack face validity;
      if implicit, author is only judge OR judges selected from a local circle of friends or acquaintances
   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

E. External validity*

8. Sample size
   0 = under 100
   1 = 100 or more

9. Number/nature of sites
   0 = single institution
   1 = multiple institutions
   2 = single geographical area or enrolled population (e.g., HMO)
   3 = multiple geographical areas or enrolled populations

10. Representativeness* in terms of age, sex, race and income
    0 = not representative
    1 = representative

* These terms are defined in detail in the text.

allow it to be generalized and applied in a broader context. Finally, a sample is said to be representative if it imitates the population under consideration [19].
Results

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. Ten articles [13,14,20–27] examined the appropriateness of procedures (mean methodology score, 8; range from 2 to 14); six articles [28–33] looked at the appropriateness of hospital use (mean score, 8; range from 3 to 12); and one article [34] 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 [24,33–34], and one examined both the misuse and underuse of hospital services for breast cancer patients [33].

Procedures – Appropriateness Table 2 provides information on the methods, e.g., sample size, population source, and results of the ten 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 inappropriate. 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 [20] analyzed the appropriateness of coronary artery bypass surgery for Medicare patients in one Western state. They found that 14% of these procedures were performed for clearly inappropriate reasons and that 30% 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. 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 [12–14,21]. If equivocal use is combined with inappropriate use, then 26, 28 and 64%, 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 [23] study. Specifically, the RAND group found that 17% of coronary angiography was inappropriate, whereas Elliott's group found that less than 3% was inappropriate. The explanation for this large difference may lie in the appropriateness judgement process. Elliott et al. specified five indications for performing the procedure and six for not performing it, while RAND specified 300 indications.
Table 2
Appropriateness of procedures, hospital use, and office care

<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>
<th>Reference number</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>
<td>20</td>
</tr>
<tr>
<td>Coronary angiography</td>
<td>2. 1677</td>
<td>3 regions</td>
<td>Benefit-risk</td>
<td>Overuse</td>
<td>17%</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Carotid endarterectomy</td>
<td>3. 1302</td>
<td>3 regions</td>
<td>Benefit-risk</td>
<td>Overuse</td>
<td>32%</td>
<td>13</td>
<td>21</td>
</tr>
<tr>
<td>Upper GI endoscopy</td>
<td>4. 1585</td>
<td>3 regions</td>
<td>Benefit-risk</td>
<td>Overuse</td>
<td>17%</td>
<td>13</td>
<td>14</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>
<td>22</td>
</tr>
<tr>
<td>Multiple Procedures: Appendectomy, Cataract removal, Cholecystectomy, Coronary angiography, Uterine dilation &amp; curettage, Hiatal hernia repair, Abdominal hysterectomy, Vaginal hysterectomy, Lumbar disk excision, Meniscectomy, Tonsillectomy, Adenoidectomy</td>
<td>6. 4850</td>
<td>1 state</td>
<td>Implicit</td>
<td>Overuse</td>
<td>3%</td>
<td>6</td>
<td>23</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>
<td>24</td>
</tr>
<tr>
<td>Preoperative lab screening</td>
<td>8. 2000</td>
<td>1 site</td>
<td>Implicit</td>
<td>Overuse</td>
<td>60%</td>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td>Cardiac pacemaker</td>
<td>9. 382</td>
<td>30 hospitals</td>
<td>Implicit</td>
<td>Overuse</td>
<td>20%</td>
<td>3</td>
<td>26</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>
<td>27</td>
</tr>
<tr>
<td>Hospital use</td>
<td>#</td>
<td>Setting</td>
<td>Risk</td>
<td>Use</td>
<td>%</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>---</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>
<td></td>
</tr>
<tr>
<td>1. 1132</td>
<td></td>
<td>6 sites in 4 regions</td>
<td>Benefit-risk</td>
<td>Overuse</td>
<td>10-35%</td>
<td>12</td>
<td>28</td>
</tr>
<tr>
<td>2. All Medicare &amp; Medicaid</td>
<td></td>
<td>25 hospitals in 4 regions</td>
<td>Implicit</td>
<td>Overuse</td>
<td>12-28% hospital admissions</td>
<td>6</td>
<td>29</td>
</tr>
<tr>
<td>3. 6574 Medicare</td>
<td></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>30</td>
</tr>
<tr>
<td>4. 200</td>
<td></td>
<td>1 site</td>
<td>Benefit-risk</td>
<td>Overuse</td>
<td>33% fee-for-service inappropriate; 27% HMO inappropriate</td>
<td>8</td>
<td>31</td>
</tr>
<tr>
<td>5. 1456 Medicare</td>
<td></td>
<td>4 hospitals</td>
<td>Implicit</td>
<td>Overuse</td>
<td>6%</td>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td>6. 374 patients</td>
<td></td>
<td>7 hospitals</td>
<td>Benefit-risk</td>
<td>Underuse, misuse</td>
<td>4.17%</td>
<td>11</td>
<td>33</td>
</tr>
<tr>
<td>Hospital care for breast cancer patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Office visits For five common conditions</td>
<td></td>
<td>Population-based</td>
<td>Implicit</td>
<td>Underuse</td>
<td>26-80%</td>
<td>7</td>
<td>34</td>
</tr>
</tbody>
</table>
The hip arthroplasty study represents an indirect attempt to define underuse. This study began by assuming that all hip replacements in one Minnesota county served by the Mayo Clinic were appropriate [24]. 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% 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 be increased if data from them could be linked to information on the amount of resources consumed by such procedures. Except for a few procedures, this unfortunately cannot be done because of limited data on appropriateness.

Each year, one in 500 Americans (mostly elderly) receive a pacemaker at an approximate annual expense of $2 billion [26]. If, as Greenspan et al. [26] found, 20% of them are inappropriate, then almost half a billion dollars a year could be saved if its inappropriate use were eliminated, and if other medical expenses were not incurred by those people who did not receive a pacemaker.

In 1985, an estimated 200,000 coronary artery bypass surgeries were performed in the United States, with costs ranging from $18,000 to $59,000 per procedure [35]. Again, using some simple naive assumptions, and eliminating both equivocal and unnecessary procedures, 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 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).

**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 clinical criteria, the Appropriateness Evaluation Protocol (AEP) developed by Gertman and Restuccia, to judge appropriateness [28,29,31–32]. 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 [36]. 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.

The articles we reviewed found a substantial amount of hospital overuse. One study found that one in five admissions and one in four days of care was inappropriate [29]. Another study found that 22% of the patients admitted to the hospital had at least one inappropriate hospital day [32], and a third study demonstrated that 23% of admissions and 34% of hospital days were inappropriate [28].

Three studies examined factors that might reduce hospital overuse. Two of them used concurrent physician feedback of the appropriateness judgement as an intervention to reduce inappropriate hospital care [29–30]. 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 [31].
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. [31] found that the overall
rate of hospital admissions judged to be discretionary was 32% in the fee-for-
service system, and 27% 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. [33], 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% of patients 30 to 69, versus 17% 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% inappropriate rates of hospital use cited in these studies are too high to be
ignored.

**Ambulatory Care – Appropriateness** We found one article that studied the
appropriateness of ambulatory care in the elderly (Table 2). Heller et al. [34]
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% ‘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 fac-
tors (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
due to 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 [17,37–54] that presented data on the appropriateness of medication use in the elderly. 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 four 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. 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.

Medications – Appropriateness

Table 3 gives 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 inappropriateness’. 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.

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 only studied three of the eight classes [55] that comprise 50% of medication use in the United States. Just over 40% of the studies collected data on the medication use of non-institutionalized, ambulatory persons. Five studies examined hospitalized populations and two examined those in nursing homes.

Twelve of the 19 studies examined overuse of medication. For example, studies of psychoactive drugs showed that between 7% and 51% of sedatives, hypnotics, antidepressants and antipsychotics are overused. Antianxiety medication was the most overused. Another study found that 22% of emergency room antibiotic prescriptions were inappropriately overused [46].

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
Table 3
Appropriateness of medication use by medication class

<table>
<thead>
<tr>
<th>Medication class</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>
<th>Reference number</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. 573</td>
<td>Hospitalized</td>
<td>Implicit</td>
<td>Overall</td>
<td>5.3% (on admission)</td>
<td>4</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>2. 389,981(2)</td>
<td>Outpatient</td>
<td>Benefit-risk</td>
<td>Misuse/overuse</td>
<td>33% overuse; 18% misuse</td>
<td>4</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>3. 30,000(1)</td>
<td>Outpatient</td>
<td>Implicit</td>
<td>Misuse</td>
<td>50% misuse</td>
<td>8</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>4. 948(1)</td>
<td>Outpatient</td>
<td>Benefit-risk</td>
<td>Overuse</td>
<td>37% overuse</td>
<td>9</td>
<td>40</td>
</tr>
<tr>
<td>Psycho-actives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. 5,902</td>
<td>Nursing home</td>
<td>Implicit</td>
<td>Misuse &amp; overuse</td>
<td>Large misuse &amp; overuse</td>
<td>5</td>
<td>41</td>
</tr>
<tr>
<td></td>
<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>42</td>
</tr>
<tr>
<td></td>
<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>43</td>
</tr>
<tr>
<td></td>
<td>8. 45</td>
<td>Outpatient</td>
<td>Implicit</td>
<td>Misuse &amp; overuse</td>
<td>11% misuse; 51% overuse</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>Anti-infectives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9. 540</td>
<td>Nursing home</td>
<td>Implicit</td>
<td>Underuse</td>
<td>67% underuse</td>
<td>3</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>10. Large</td>
<td>Theoretical</td>
<td>Benefit-cost</td>
<td>Underuse</td>
<td>65%-80% underuse</td>
<td>10</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>11. 1,185</td>
<td>Emergency room</td>
<td>Implicit</td>
<td>Underuse &amp; overuse</td>
<td>22% inappropriate use, mostly due to overuse</td>
<td>3</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>12. 1,021</td>
<td>Hospitalized</td>
<td>Implicit</td>
<td>Misuse</td>
<td>69% misuse of prophylactic antibiotics</td>
<td>8</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>13. 267</td>
<td>Hospitalized</td>
<td>Implicit</td>
<td>Misuse</td>
<td>74% misuse of all prophylactic courses</td>
<td>8</td>
<td>48</td>
</tr>
<tr>
<td></td>
<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>49</td>
</tr>
<tr>
<td></td>
<td>15. 1,900</td>
<td>Outpatient</td>
<td>Implicit</td>
<td>Misuse</td>
<td>44% misuse in elderly</td>
<td>7</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>16. Not stated</td>
<td>Hospitalized</td>
<td>Benefit-cost</td>
<td>Misuse</td>
<td>60-90% misuse</td>
<td>7</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>17. 78(1)</td>
<td>Hospitalized</td>
<td>Implicit</td>
<td>Misuse &amp; overuse</td>
<td>17% overuse; 37% misuse</td>
<td>3</td>
<td>52</td>
</tr>
<tr>
<td>GI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18. 51</td>
<td>Outpatient</td>
<td>Benefit-risk</td>
<td>Overuse &amp; misuse</td>
<td>35% overuse; 31% misuse</td>
<td>4</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>19. 103(1)</td>
<td>Outpatient</td>
<td>Implicit</td>
<td>Misuse &amp; overuse</td>
<td>25% overuse; 71% misuse</td>
<td>4</td>
<td>54</td>
</tr>
<tr>
<td>Anti-Arthritis</td>
<td>No literature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac</td>
<td>No literature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diuretics</td>
<td>No literature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cold/Cough</td>
<td>No literature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analgesics</td>
<td>No literature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The population size is the number of people using medications, except where otherwise noted.
(1)= Number of medications.
(2)= Number of pharmacy sites.
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 receive antibiotics, the authors determined whether the failure to treat was appropriate). 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% of cases, because clinicians failed to develop an adequate treatment plan [42]. In two examples of misuse, cimetidine was prescribed for the wrong duration of time [53–54]. Two studies examined the misuse of antibiotics when used as prophylaxis for surgery. One [48] found that antibiotic prophylaxis was given for more than two days in 93% 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 [50].

To summarize our results, we compiled the findings from the ten best papers (those that scored 6 or above, the mean of all scores). The data from this compilation suggest that 7 to 37% of medication is overused in outpatients, and 44 to 90% 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 comprise over 50% 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

<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>Psychoactives</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>
</tbody>
</table>

* The reader should not multiply the numbers in the second column by those in the fourth column and conclude that this is the amount that could be saved by eliminating overuse. Savings will occur, of course, at the margin (in economic jargon).
prescribed correctly, and that we are overusing and misusing drugs. We know even less about the rate of underuse of medication.

Discussion

This article covers 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.

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 specified both a sampling frame to identify relevant literature, and then developed a rating scale by which to judge the 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 whether 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 (versus implicit criteria), does the conclusion about appropriateness of care change? If data are collected prospectively (as opposed to retrospectively), do appropriateness results change?

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 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 services to the needs of elderly patients. This mismatch occurs both in terms of overuse and underuse.

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 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 result in the freeing up of 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.

Acknowledgements


References

10 Ware, J.E. Jr., Rogers, W.H., Davies, A.R. et al., Comparison of health outcomes at a health maintenance organization with those of Fee-for-service care. Lancet ii (1986) 1017–1022.