Agreement Between Drugs-to-Avoid Criteria and Expert Assessments of Problematic Prescribing

Michael A. Steinman, MD; Gary E. Rosenthal, MD; C. Seth Landefeld, MD; Daniel Bertenthal, MPH; Peter J. Kaboli, MD

**Background:** Drugs-to-avoid criteria are commonly used to evaluate prescribing quality in elderly persons. However, few studies have evaluated the concordance between these criteria and individualized patient assessments as measures of problem prescribing.

**Methods:** We used data on 256 outpatients from the Iowa City VA Medical Center who were 65 years or older and taking 5 or more medications. After a comprehensive patient interview, a study team composed of a physician and a pharmacist recommended that certain drugs be discontinued, substituted, or reduced in dose. We evaluated the degree to which drugs considered potentially inappropriate by the drugs-to-avoid criteria of Beers et al and Zhan et al (hereinafter, Beers criteria and Zhan criteria) were also considered problematic by the study team, and vice versa.

**Results:** In the study cohort, 256 patients were using 3678 medications. The physician-pharmacist team identified 563 drugs (15%) as problematic, while 214 drugs (6%) were flagged as potentially inappropriate by the Beers criteria and 91 drugs (2.5%) were flagged as potentially inappropriate using the Zhan criteria. The k statistics for concordance between drugs-to-avoid criteria and expert assessments were 0.10 to 0.14, indicating slight agreement between these measures. Sixty-one percent of drugs identified as potentially inappropriate by the Beers criteria and 49% of drugs flagged by the Zhan criteria were not judged to be problematic by the expert reviewers. Correspondence between drugs-to-avoid criteria and expert assessment varied widely across different types of drugs.

**Conclusions:** Drugs-to-avoid criteria have limited power to differentiate between drugs and patients with and without prescribing problems identified on individualized expert review. Although these criteria are useful as guides for initial prescribing decisions, they are insufficiently accurate to use as stand-alone measures of prescribing quality.

Arch Intern Med. 2009;169(14):1326-1332

**DRUGS-TO-AVOID CRITERIA** are lists of drugs considered potentially inappropriate for elderly persons owing to adverse effects, limited effectiveness, or both. These criteria are commonly used as markers of prescribing problems for elderly patients in research and the practice of quality measurement.1,6 For example, the Centers for Medicare and Medicaid Services (CMS) mandates the use of a version of the criteria of Beers et colleagues (hereinafter, Beers criteria7) in nursing homes,8 and the National Committee for Quality Assurance uses a version of the criteria of Zhan et al1 (hereinafter, Zhan criteria) to compare the quality of US health plans.3

Despite the widespread use of drugs-to-avoid criteria, evidence of their validity as markers of prescribing quality for elderly patients is mixed. The most commonly used criteria were developed by expert panels, and there is substantial disagreement about which drugs should be included on these lists.1,9-11 For another marker of validity—the ability of these criteria to predict adverse outcomes—results of observational studies have been inconsistent.2,12-16 Interpretation of these outcomes studies is further complicated by the difficulty of isolating the impact of implicated drugs on clinical outcomes independent of characteristics of patients and concurrent therapies they received. Finally, other work17 has suggested that drugs-to-avoid criteria medications account for only a small fraction of adverse drug events.

These mixed results highlight the need to better understand the accuracy of drugs-to-avoid criteria as markers of prescribing quality. However, there are little empirical data on the extent to which drugs considered potentially inappropriate by these criteria are in fact inappropriate when reviewed in the context of case histories of actual patients. In this study, we compared 2 commonly used drugs-to-avoid criteria with individualized expert assess-
ment of patients' medications in a cohort of over 250 elderly veterans. In doing so, we focused on whether drugs considered inappropriate by the Beers and Zhan criteria were also considered inappropriate when evaluated by individualized expert review.

**METHODS**

**DATA SOURCE**

We used data from the Enhanced Pharmacy Outpatient Clinic (EPOC) trial. This randomized controlled trial evaluated the impact of a specialized medication review clinic on prescribing and patient outcomes among older veterans in the outpatient clinics of the Iowa City VA Medical Center. Eligibility criteria for participation in the trial included an age of 65 years or older and use of 5 or more medications.

Patients in the intervention arm were evaluated in a medication review clinic. During the baseline visit, a study pharmacist with expertise in geriatric pharmacotherapy conducted an in-depth interview about the patient's medication use and adverse effects of their medications. The medication list generated from this interview included all drugs, supplements, and herbal preparations currently used by the patient, including prescription and nonprescription drugs from both VA and non-VA sources. Toward the end of the visit, a physician with expertise in prescribing for elderly patients conferred with the pharmacist and patient, after which the physician-pharmacist team generated a consensus list of recommendations that was delivered to the patient's primary care physician. The physician-pharmacist team identified problems using implicit review and then categorized the problems and recommended responses using a process with substantial to excellent interrater reliability (κ statistic, 0.64-0.85). These recommendations included suggestions about drugs that should be discontinued, substituted with a different drug, or prescribed at different doses, as well as suggestions about initiating new drugs that may benefit the patient. For example, one interview identified a patient taking a calcium-channel blocker who had developed lower extremity edema, and the team recommended that the drug be stopped. In another case, an older patient taking a highly anticholinergic drug may have reported no adverse effects and good effectiveness of the drug, in which case the team did not recommend a change in therapy. All recommendations were recorded and categorized in the study database and assigned a priority level of high, medium, or low. It is worth noting that the expert reviewers did not explicitly consult the Beers criteria or other such lists in making their recommendations. Nonetheless, as experts in prescribing for elderly patients, they were aware of these criteria and likely incorporated their principles into their clinical recommendations.

Other data, including basic demographic information, medical history, and health care utilization, were collected from patients at the enrollment interview and through review of medical records. Of 258 patients enrolled in the intervention arm, the medication list and expert recommendations were not available for 2. The remaining 256 patients comprised our study population.

**MEASURES**

We separately evaluated 2 drugs-to-avoid criteria commonly used in research and quality assessment. We used data from the baseline medication list and medical history to encode each patient's drugs as meeting or not meeting each of the criteria listed below.

**Beers Criteria**

The Beers criteria include a list of drugs that are considered inappropriate for all elderly patients (eg, propoxyphene hydrochloride), drugs that should not be prescribed above certain doses (eg, ferrous sulfate at >325 mg/d), and drug-disease and drug-drug combinations to avoid (eg, anticholinergic medications in patients with bladder outflow obstruction). Each criterion on the list is classified as high or low severity. We evaluated all of the criteria specified in the most recent update (2003), including consideration of drug doses, drug-disease interactions, and drug-drug combinations.

**Zhan Criteria**

Based on the 1997 version of the Beers criteria, the Zhan criteria focus only on drugs that should generally be avoided in elderly patients, without consideration of drug dosages, drug-disease interactions, or drug-drug combinations. The Zhan criteria categorize drugs into 1 of 3 categories: drugs that should always be avoided (eg, meperidine hydrochloride), drugs that are rarely appropriate (eg, diazepam), and drugs that are sometimes appropriate but often misused (eg, amitriptyline hydrochloride).

**Expert Review**

We used the study database to identify all drugs from the baseline interview that were recommended by the physician-pharmacist team to be discontinued, substituted with another drug, or prescribed at a lower dose. We considered any of these recommendations to be a prescribing problem as judged by the expert reviewer.

**STATISTICAL ANALYSIS**

We attempted to match all potentially inappropriate medications (PIMs) identified by the Beers and Zhan criteria with recommendations made by the expert team. In cases where a PIM was recommended to be discontinued, substituted with another drug, or to have its dose lowered, we considered the expert assessment to be concordant with the drugs-to-avoid criteria. All analyses were performed at the level of the drug (n=3678). In addition, we repeated our analyses at the level of the patient (n=256), whereby any positive result on the Beers criteria, Zhan criteria, or expert review would identify that patient as having "problematic prescribing." Because 208 of 256 patients had at least 1 drug change recommendation, for this analysis we considered patients to have a prescribing problem if expert review yielded a high-priority recommendation (n=166).

We approached our analyses from 2 complementary perspectives. First, we evaluated the concordance between drugs-to-avoid criteria and individualized expert review using κ statistics, which provide a measure of agreement between separate ratings of the same construct—in this case, 2 methods of determining whether a drug was problematic or not problematic—beyond the agreement that would be expected by chance. However, because drugs-to-avoid criteria and individualized expert assessments are designed to measure different aspects of prescribing quality, one would not expect a high κ value even if both evaluations perfectly captured the elements of prescribing quality they attempt to measure. Thus, we used a second approach whereby we considered the expert assessment a de facto reference standard and compared the sensitivity and specificity of drugs-to-avoid criteria in comparison with this standard. Such expert assessments are not universally accepted as a criterion standard for defining prescribing quality, in part be-
perform in a clinically individualized, real-life setting. 

Son to improve understanding of how drugs-to-avoid criteria by a thoughtful clinician make them a useful comparison only and 1243 over-the-counter drugs, including vitamins, minerals, and herbal preparations. Patients were predominantly white and male (Table 1). Of the 3678 drugs taken by study participants (n=3678b), drugs per participant, mean±SD 14.4±5.0. Prescription-only "regular" drugs 8.7±3.8. Prescription-only other drugs 0.8±1.1. Nonprescription "regular" drugs 2.3±1.4. Nonprescription other drugs 2.6±2.1. Problem drugs per expert review 563 (15.3). By priority, high 308 (8.4). Intermediate 190 (5.2). Low 65 (1.8). By recommendation, discontinue drug 379 (10.3). Substitute with another drug 105 (2.9). Reduce dose of drug 79 (2.2). Problem drugs per Beers criteria 214 (5.8). "High severity" 156 (4.2). "Low severity" 58 (1.6). Problem drugs per Zhan criteria 91 (2.5). "Always avoid" 1 (<0.1). "Rarely appropriate" 16 (0.4). "Sometimes appropriate but often misused" 74 (2.0). 

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study participants (n=256)</td>
<td></td>
</tr>
<tr>
<td>Age, mean±SD, y</td>
<td>73.5±5.1</td>
</tr>
<tr>
<td>Male sex</td>
<td>249 (97)</td>
</tr>
<tr>
<td>White race</td>
<td>252 (98)</td>
</tr>
<tr>
<td>Clinical conditions</td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>213 (83)</td>
</tr>
<tr>
<td>Ischemic heart disease</td>
<td>138 (54)</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>105 (41)</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td>91 (36)</td>
</tr>
<tr>
<td>Depression</td>
<td>47 (18)</td>
</tr>
<tr>
<td>Drugs taken by study participants (n=3678)b</td>
<td></td>
</tr>
<tr>
<td>Drugs per participant, mean±SD</td>
<td>14.4±5.0</td>
</tr>
<tr>
<td>Prescription-only &quot;regular&quot; drugs</td>
<td>8.7±3.8</td>
</tr>
<tr>
<td>Prescription-only other drugs</td>
<td>0.8±1.1</td>
</tr>
<tr>
<td>Nonprescription &quot;regular&quot; drugs</td>
<td>2.3±1.4</td>
</tr>
<tr>
<td>Nonprescription other drugs</td>
<td>2.6±2.1</td>
</tr>
<tr>
<td>Problem drugs per expert review</td>
<td>563 (15.3)</td>
</tr>
<tr>
<td>By priority</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>308 (8.4)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>190 (5.2)</td>
</tr>
<tr>
<td>Low</td>
<td>65 (1.8)</td>
</tr>
<tr>
<td>By recommendation</td>
<td></td>
</tr>
<tr>
<td>Discontinue drug</td>
<td>379 (10.3)</td>
</tr>
<tr>
<td>Substitute with another drug</td>
<td>105 (2.9)</td>
</tr>
<tr>
<td>Reduce dose of drug</td>
<td>79 (2.2)</td>
</tr>
<tr>
<td>Problem drugs per Beers criteria</td>
<td>214 (5.8)</td>
</tr>
<tr>
<td>&quot;High severity&quot;</td>
<td>156 (4.2)</td>
</tr>
<tr>
<td>&quot;Low severity&quot;</td>
<td>58 (1.6)</td>
</tr>
<tr>
<td>Problem drugs per Zhan criteria</td>
<td>91 (2.5)</td>
</tr>
<tr>
<td>&quot;Always avoid&quot;</td>
<td>1 (&lt;0.1)</td>
</tr>
<tr>
<td>&quot;Rarely appropriate&quot;</td>
<td>16 (0.4)</td>
</tr>
<tr>
<td>&quot;Sometimes appropriate but often misused&quot;</td>
<td>74 (2.0)</td>
</tr>
</tbody>
</table>

4 Data are given as number (percentage) except where noted. 

b "Regular" drugs refers to systemically active (eg, oral) drugs other than vitamins, minerals, and herbal preparations. "Problem drugs per expert review" includes drugs recommended to be discontinued, substituted, or reduced in dose. "Other drugs" indicates topical drugs, vitamins, minerals, and herbal preparations. Nonprescription drugs are drugs available over-the-counter, regardless of whether the drug was recommended or ordered by a physician. 

c See study by Zhan et al.1 

d See study by Fick et al.7 

e See study by Zhan et al.1 

cause reviewers may reasonably disagree in their assessments of a given patient’s medications and because there is limited evidence of these reviews’ impact on clinical outcomes.21-23 None- of a given patient’s medications and because there is limited cause reviewers may reasonably disagree in their assessments of how drugs-to-avoid criteria perform in a clinically individualized, real-life setting.21,24,25

This research was approved by the institutional review boards at the University of California, San Francisco, and the University of Iowa, and by the Research and Development Committees of the San Francisco and Iowa City VA Medical Centers. The sponsors had no control over the study question, analyses, or decision to publish this article.

RESULTS

The study sample comprised 256 patients taking 3678 drugs, including 2425 drugs available through prescription only and 1243 over-the-counter drugs, including vitamins, minerals, and herbal preparations. Patients were predominantly white and male (Table 1). Of the 3678 medications assessed by the expert physician-pharmacist team, 563 (15.3%) were considered problematic as reflected by a recommendation to discontinue the drug, substitute it with another drug, or reduce the dose (Table 1). The Beers criteria identified 214 of 3678 drugs (5.8%) as potentially inappropriate, and the Zhan criteria identified 91 drugs (2.5%) as potentially inappropriate. The most common classes of drugs identified by the Beers and Zhan criteria are shown in Table 2.

The Figure shows the correspondence between the expert recommendations and the Beers and Zhan criteria. The κ statistic was 0.14 (95% confidence interval [CI], 0.10-0.18) for the Beers criteria and 0.10 (95% CI, 0.07-0.14) for the Zhan criteria, indicating “slight” agreement between the drugs-to-avoid criteria and individualized expert review beyond what would be expected by chance (Figure, A and B). Among 214 drugs meeting 1 or more of the Beers criteria, 83 (39%) were considered problematic by expert review. Among 563 drugs considered to be problematic by expert review, 83 (15%) were considered to be problematic according to the Beers criteria. Results for the Zhan criteria followed a generally similar pattern: 46 of 91 drugs (51%) flagged by the Zhan criteria were deemed to be problematic by expert review, while 46 of 563 drugs (8%) flagged by expert review were deemed to be problematic according to the Zhan criteria.

Expert reviewers cited a variety of reasons for recommending discontinuation, substitution, or dose reduction of the 480 drugs that they but neither the Beers nor Zhan criteria identified as problematic. Among these 480 drugs, 61 (13%) were flagged as causing actual adverse drug reactions, and an additional 111 (23%) were flagged as causing potential adverse drug reactions. In addition, 138 drugs (29%) had problems relating to indications (eg, drugs that lacked indications or provided suboptimal treatment for the condition of interest); 105 (22%) had problems with effectiveness (eg, minimal or no evidence of therapeutic effectiveness); 53 (11%) had problems with inappropriate dose, schedule, or therapeutic duplication; and 12 (3%) had miscellaneous other problems. Among the 83 drugs identified as problematic by both the experts and the Beers or Zhan criteria, 49 (59%) were flagged by experts on the basis of real or potential adverse drug reactions, and 22 (27%) were flagged on the basis of lacking indications or providing suboptimal treatment for the condition of interest.

The correspondence between drugs-to-avoid criteria and expert assessment varied across different types of drugs (Table 2). For example, nearly all of the tricyclic antidepressants identified as problematic by the Beers and Zhan criteria were also implicated by the expert assessment. In contrast, there was an almost complete lack of overlap in assessments of muscle relaxants. Among 10 cases of cyclobenzaprine hydrochloride use identified by the Beers and Zhan criteria, only 1 was rated as problematic by the expert team. However, the expert team recommended changes for 2 of 4 prescriptions of the muscle relaxant baclofen (which is not included in the Beers and Zhan criteria).

Our next analyses focused on results at the level of the patient (Figure, C and D). Overall, 136 patients (53%) were taking at least 1 drug identified as problematic by the Beers criteria, and 71 patients (28%) were taking at least 1 drug
identified as problematic by the Zhan criteria. The \( \kappa \) statistic for concordance with expert review was 0.14 (95% CI, 0.02-0.26) for the Beers criteria and 0.17 (95% CI, 0.08-0.25) for the Zhan criteria, indicating that most of the observed agreement between drugs-to-avoid criteria and expert review could be attributed to chance alone.

Next, we assessed the performance of drugs-to-avoid criteria in a setting where we designated the expert review as a gold standard for detecting prescribing problems. In our cohort, 39% of drugs flagged by the Beers criteria and 51% of drugs flagged by the Zhan criteria were considered problematic by expert review (positive predictive value), with positive likelihood ratios of 3.5 and 5.7, respectively. When evaluated at the level of the patient, the ability of the Beers and Zhan criteria to distinguish between patients with prescribing problems vs those without fell further, with positive likelihood ratios of 1.3 and 2.5, respectively.

We conducted a number of sensitivity analyses in which we varied the thresholds for determining a drug to be problematic, including thresholds for the Beers criteria, the Zhan criteria, and the expert assessments. Most permutations yielded results similar to our main analyses (eTable, http://www.archinternmed.com). Finally, because the Beers and

Table 2. Most Common Drugs Identified by the Beers and Zhan Criteria and Expert Assessment of Those Drugs

<table>
<thead>
<tr>
<th>Type of Drug</th>
<th>Beers Criteria (214 Drugs Flagged)</th>
<th>Zhan Criteria (91 Drugs Flagged)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Drugs Meeting Criteria, No.</td>
<td>Drugs That Met Criteria and Were Deemed Problematic by Expert Team, No. (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antihistamines</td>
<td>42</td>
<td>21 (50)</td>
</tr>
<tr>
<td>Tricyclic antidepressants</td>
<td>18</td>
<td>15 (83)</td>
</tr>
<tr>
<td>Benzodiazepines</td>
<td>15</td>
<td>5 (33)</td>
</tr>
<tr>
<td>Digoxin (^b)</td>
<td>15</td>
<td>10 (67)</td>
</tr>
<tr>
<td>Urinary antispasmodics</td>
<td>12</td>
<td>6 (50)</td>
</tr>
<tr>
<td>Pseudoephedrine</td>
<td>12</td>
<td>6 (50)</td>
</tr>
<tr>
<td>Skeletal muscle relaxants</td>
<td>10</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Ticlopidine and clopidogrel</td>
<td>6</td>
<td>1 (17)</td>
</tr>
<tr>
<td>Propoxyphene</td>
<td>4</td>
<td>1 (25)</td>
</tr>
</tbody>
</table>

Abbreviation: NA, not covered by the Zhan criteria (Zhan et al\(^1\)).

\(^a\)Discrepancies between the number of problem drugs identified using the Beers criteria (see study by Fick et al\(^7\)) and Zhan criteria reflect different criteria for each drug class. For example, the Beers criteria recommend against a variety of benzodiazepines in different settings, whereas the Zhan criteria cover only 3 specific benzodiazepines.

\(^b\)Digoxin is rated as problematic by the Beers criteria if prescribed at a dose higher than 0.125 mg/d. Expert assessment included both recommendations for drug discontinuation and dose reduction.
Zhan criteria focus principally on systemically administered allopathic medications, we repeated our analyses after excluding 585 topical preparations, herbal medications, and multivitamins. Results of these analyses were similar to the main analyses.

### Table 3. Test Characteristics of Drugs-to-Avoid Criteria Compared With Individualized Expert Review

<table>
<thead>
<tr>
<th>Drug or Patient</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive PV</th>
<th>Negative PV</th>
<th>Positive Likelihood Ratio</th>
<th>Negative Likelihood Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>For 3678 individual drugs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beers criteria(^b)</td>
<td>15</td>
<td>96</td>
<td>39</td>
<td>86</td>
<td>3.5</td>
<td>0.9</td>
</tr>
<tr>
<td>Zhan criteria(^c)</td>
<td>8</td>
<td>99</td>
<td>51</td>
<td>86</td>
<td>5.7</td>
<td>0.9</td>
</tr>
<tr>
<td>For 256 patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beers criteria(^b)</td>
<td>58</td>
<td>57</td>
<td>71</td>
<td>43</td>
<td>1.3</td>
<td>0.7</td>
</tr>
<tr>
<td>Zhan criteria(^c)</td>
<td>35</td>
<td>86</td>
<td>82</td>
<td>42</td>
<td>2.5</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Abbreviation: PV, predictive value.

\(^a\) In this table, individualized expert review is the reference standard against which drugs-to-avoid criteria are evaluated. Test characteristics for individual drugs correspond to the data in Figure A and B, and test characteristics for patients (where problem prescribing is defined as use of 1 or more problematic drugs) correspond to the data in Figure C and D. Data are given as percentages except where noted.

\(^b\) See study by Fick et al.\(^7\)

\(^c\) See study by Zhan et al.\(^1\)

In this study of elderly veterans, we found substantial discordance between drug quality assessments made by drugs-to-avoid criteria and individualized expert assessments. Half or more of the drugs flagged by the Beers and Zhan criteria were not considered problematic on individualized, implicit expert review. Moreover, the Beers and Zhan criteria identified only 8% to 15% of drugs that experts judged to be problematic. Similarly discordant results were observed at the level of the patient, with limited correlation between patients taking drugs-to-avoid medications and those with prescribing problems identified on expert review.

Our finding that drugs-to-avoid criteria detected only a small fraction of prescribing problems found on individualized expert review is not surprising. Drugs-to-avoid criteria are not intended to identify all problematic drugs but rather to have high specificity and high positive predictive value; that is, to focus on a limited number of drugs for which consensus indicates that use is often (or almost always) inappropriate.\(^1,15\) However, our findings suggest suboptimal accuracy of the Beers and Zhan criteria even for this limited goal. Half or more of the drugs identified as problematic by the Beers and Zhan criteria were not judged as problematic by the expert reviewers. Although the developers of these criteria were careful to note that there may be exceptions to the judgments rendered by their criteria, these exceptions were as or more common than the rule. These findings support the claim, frequently made by physicians, that many of the drugs included in the Beers and Zhan drugs-to-avoid criteria are appropriate in selected circumstances.\(^3,9\) Of note, there is no single, universally accepted standard for defining prescribing problems, so we cannot definitively conclude that the drugs-to-avoid criteria were incorrect in every instance where they disagreed with individualized expert review. Nonetheless, to the extent that individualized drug review represents a careful, patient-oriented assessment in real-world clinical settings, our findings suggest that drugs-to-avoid criteria have limited ability to distinguish between drugs that pose a problem for patients vs those that do not.

In addition to their limitations in evaluating individual drugs, our findings suggest limited accuracy of drugs-to-avoid criteria when applied at the level of the patient (defined by the presence or absence of an offending drug on the patient’s medication list). Concordance between the Beers criteria and expert review was only slightly above that expected by chance, with the Beers criteria having almost no ability to discriminate between patients with and without prescribing problems defined by expert review (as reflected by likelihood ratios close to 1.0). The Zhan criteria had a positive likelihood ratio of 2.5, somewhat better than the Beers criteria but still reflecting a weak ability to distinguish between patients with prescribing problems vs those without problems identified on expert review.

These results follow a limited body of previous work. In a small study\(^26\) of a homeless geriatric population, a clinical pharmacist recommended drug changes for 60% of drugs flagged by the Beers criteria identified on medical record review (76% when previously discontinued drugs were excluded). In contrast, another study\(^4\) performed in nursing homes identified uneven and generally minimal changes in use of medications from a drugs-to-avoid list after the CMS implemented a policy mandating utilization review of patients taking these drugs, suggesting that most such drugs were maintained even after individualized review. Finally, in a previous report\(^27\) from the EPOC study we found low levels of interrater reliability between the Beers criteria and other commonly used measures of prescribing quality, including the Medication Appropriateness Index and use of 9 or more medications (one definition of “polypharmacy”).

Despite the limitations of drugs-to-avoid criteria, the Beers and Zhan criteria are useful when applied in a suitable context. First, these criteria may be useful for identifying prescribing problems in a retrospective review of elderly patients’ medication lists.\(^26\) This application shows promise insofar as it uses drugs-to-avoid criteria to screen drugs for individualized review rather than using the criteria as the final arbiter of appropriateness.\(^8\) Second, drugs-to-avoid criteria may be particularly valuable when ap-
medications. Our study demonstrates substantial de-
shortcomings when these criteria are used for this purpose. In
particular, we found that half or more of the quality “prob-
lems” identified by the criteria may in fact not have been
problems. The ambiguity of quality judgments made by
drugs-to-avoid criteria are further amplified when com-
paring care across physicians or institutions. Given that the
appropriateness of these drugs may vary substantially across
different clinical settings and that the number of medica-
tions a patient receives is strongly linked to the presence
doctors included in the Beers and Zhan criteria, comparis-
ons of prescribing quality using drugs-to-avoid criteria may
be particularly challenging when patients’ clinical sce-
narios, level of illness burden, and medication use vary be-
tween institutions or physicians.3,9,10,13

Our results should be interpreted in the context of our
study design and limitations of our measures. First, pa-
tients were recruited from a single VA medical center and
were taking a minimum of 5 medications. Second, the
expert pharmacist reviews are an imperfect measure of
prescribing quality, and different experts may give dif-
ferent assessments of prescribing appropriateness. (This
study did not conduct dual independent ratings of ap-
propriateness for each patient, although the ultimate de-
cision about prescribing recommendations were made by
consensus by an expert pharmacist and physician, thus
limiting the impact of any one rater to influence the re-
results.) Thus, our expert reviews should not be consid-
ered a criterion standard of prescribing quality, and fur-
ther studies are needed to confirm our findings in different
care settings and with different expert raters. Third, the
recommendations generated by the study’s expert raters
reflected the individual clinical circumstances of the pa-
tient. Thus, our results should be interpreted as evalu-
ating drugs-to-avoid criteria against real-world clinical

Accepted for Publication: April 20, 2009.
Correspondence: Michael A. Steinman, MD, San
Francisco VAMC, Box 181G, 4150 Clement St, San Francisco,
CA 94121 (mike.steinman@ucsf.edu).

Author Contributions: Dr Steinman had full access to all
the data in the study and takes responsibility for the integ-
rely on the integrity of the data and the accuracy of the data analy-
ysis. Study concept and design: Steinman, Rosenthal, Lan-
defeld, and Kaboli. Acquisition of data: Steinman, Rosen-
thal, Bertenthal, and Kaboli. Analysis and interpretation of
data: Steinman, Rosenthal, Bertenthal, and Kaboli. Drafting of
the manuscript: Steinman and Kaboli. Critical revision of
the manuscript for important intellectual content: Rosen-
thal, Landefeld, Bertenthal, and Kaboli. Statistical analy-
sis: Steinman and Rosenthal. Obtained funding: Rosen-
thal and Kaboli. Administrative, technical, and material
support: Steinman, Bertenthal, and Kaboli. Study supervi-
sion: Steinman and Rosenthal.

Financial Disclosure: None reported.

Funding/Support: This study was supported by the Health
Services Research and Development Service, Depart-
ment of Veterans Affairs through an investigator-
initiated research award (SAF98-152) to Dr Rosenthal;
Research Career Development awards to Drs Steinman
and Kaboli (RCD 01-013 and RCD 03-033, respec-
tively); by a Career Development Award from the Na-
tional Institute on Aging (1K23AG030999) to Dr Stein-
man; by the Center for Research in the Implementation
of Innovative Strategies in Practice (HFP 04-149) at the
Iowa City VA Medical Center to Drs Rosenthal and Kaboli;
Agency for Healthcare Research and Quality Centers for
Education and Research on Therapeutics cooperative
agreement 5 U18 HS016094; and support from the Health
Service Research and Development Research Enhance-
ment Award Program at the San Francisco VA Medical
Center to Mr Bertenthal. Additional support was pro-
duced by grants from the National Institute on Aging (AG
00912 and AG 10418) and the John A. Hartford Found-
dation Inc to Dr Landefeld.

Role of the Sponsor: None of the sponsors had any role
in the study design, methods, analyses, and interpreta-
tion; in the preparation of the manuscript; or in the de-
cision to submit it for publication.

Disclaimer: The views expressed in this article are those
of the authors and do not necessarily reflect the posi-
tion or policy of the Department of Veterans Affairs.
In 2006, nearly 20% of adults 65 years or older in the United States took at least 10 medications, nearly 60% took 5 or more medications, and over 90% took at least 1 medication a week. In 2007, total Medicare spending on outpatient prescription drugs (part D) increased 17.5% to $47.6 billion. The number of drugs used by older adults will likely further increase owing to an aging population, the development of new prescription medications, the transition of prescription medications to over-the-counter availability, and the increasing use of drugs for chemoprevention.

Physicians and pharmacists, health services researchers and pharmacologists, policy makers, and payers have been struggling to improve the safety and effectiveness of prescribing for older adults for decades. Most of this effort has focused on identifying medications that might be associated with elevated risks and reduced benefits in older adults and reducing the use of these medications. In 1991, Beers et al. introduced criteria to help researchers evaluate prescribing quality in nursing homes (hereinafter, Beers criteria). In 2001, Zhan et al. published a subset of the criteria by Beers et al for evaluating prescribing quality in outpatient populations (hereinafter, Zhan criteria).

Other measures of potentially inappropriate medication use in older adults have been used in Canada, and new measures have recently been developed specifically for French and Irish populations.