
Drug Use and Prescribing Problems in Four State Medicaid Programs

Bruce Stuart, Ph.D., Becky A. Briesacher, M.A., Frank Ahern, Ph.D., David Kidder, Ph.D., Christopher Zacker, Ph.D., Gary Erwin, Pharm.D., Daniel Gildea, M.S., and Cheryl Fahlman, R.Ph., M.B.A.

In this article the authors present population-level prevalence rates for 61 specific drug-related problems occurring in three State Medicaid programs (Maryland, Iowa, and Washington) from 1989 through 1996 and a fourth (Georgia) from 1994 through 1996. The findings represent the first application of a consistent drug utilization review (DUR) screener program to Medicaid data across States. The study finds major differences in DUR failure rates among the four States with the lowest rates in Georgia and the highest in Washington. Only Iowa showed any population-level reduction in DUR failure rates during the study period, however, rates for community-dwelling elderly fell in most States.

BACKGROUND

Recent studies suggest that medication-related problems are a serious and growing public health issue in the United States. In a widely quoted article, Johnson and Bootman (1995) estimate that drug-related morbidity and mortality account for nearly \$77 billion per year in added health care expenditures. Bates, Spell, and Cullen (1997) calculate the annual cost of treating

adverse drug reactions (ADRs) among hospitalized patients nationwide at more than \$4 billion. Bootman, Harrison, and Cox (1997) place a similar \$4 billion price tag on the cost of treating medication-related problems in nursing homes. A meta-analysis of 39 prospective studies conducted over the past 20 years of the incidence of ADRs in hospitals led Lazarou, Pomeranz, and Corey (1998) to conclude that 2.2 million patients suffered serious ADRs in 1994, and 106,000 died because of them—making ADRs the fifth-leading cause of death for hospitalized patients in that year. A recent Alliance for Aging Research Report (1998) lists medication-related problems as the fifth-most-costly “disease” affecting the elderly, at \$84.6 billion, following diabetes (\$92 billion) but ahead of musculoskeletal conditions (\$64.8 billion).

These are staggering statistics. Indeed, the numbers are so huge that they engender skepticism about the true extent of the problem. Adding to the skepticism is the fact that none of the estimates just quoted is based on population-level data. The Johnson and Bootman (1995) and Bootman, Harrison, and Cox (1997) studies are based on the judgments of expert clinical panels combined with unit-cost data from Medicare payment rates. The Bates, Spell, and Cullen (1997) and Lazarou, Pomeranz, and Corey (1998) articles extrapolate ADR rates from small samples of mainly teaching hospitals to the country at large with no adjustments for differences in hospital or patient characteristics. The Alliance for Aging Research

Bruce Stuart, Becky Briesacher, and Cheryl Fahlman are with the University of Maryland School of Pharmacy; Frank Ahern is with the Pennsylvania State University; David Kidder is with Abt Associates, Inc.; Christopher Zacker is with Novartis Pharmaceutical Corporation; Gary Erwin is with Omnicare, Inc.; and Daniel Gildea is with Jen Associates, Inc. This research was performed under Health Care Financing Administration (HCFA) Contract Number 500-93-0002. The views and opinions expressed in this article are those of the authors and do not necessarily reflect the views of the University of Maryland School of Pharmacy, the Pennsylvania State University, Abt Associates, Inc., Novartis Pharmaceutical Corporation, Omnicare, Inc., Jen Associates, Inc., or HCFA.

estimate is apparently derived from Johnson and Bootman's figures but attributes the entire cost of ADRs to persons age 65 or over.

The lack of credible, population-based data on medication-related problems makes it difficult to formulate appropriate policies to combat them. One data source that could potentially provide such information has been largely overlooked. All State Medicaid programs are required by the Omnibus Budget Reconciliation Act of 1990 (OBRA 90) to implement DUR programs. The vast majority of Medicaid programs have computerized their review process to check each prescription before it is filled for an array of potential problems. These problems include drug-drug interactions, age and pregnancy alerts, proper dosage, early refills, and many other concerns. Because automated DUR programs maintain central repositories of prescription claims information, they can be programmed to produce exceptions and drug-problem reports at various levels of aggregation. But there are serious drawbacks to relying on such systems for population-level data on drug-related problems. Most systems are operated by proprietary vendors whose criteria of appropriate drug use or computer algorithms are not in the public domain. Furthermore, these systems have not been independently validated or cross-validated with one another. Lastly, the DUR vendors do not publish their findings.

These concerns are not intrinsic to computerized DUR screening. Rather, the issue is a lack of transparency by which the criteria and computer algorithms are made publicly available for scientific scrutiny and validation. The research described in this article is based on a DUR screening system that offers an open architecture of both the DUR criteria and programming conventions. Known as the Pennsylvania

State University (PSU) screener, this is an expert system developed by PSU under a cooperative agreement with HCFA (Stuart, Ahern, and Coulson, 1994). The present study used this screening program to compute population-level prevalence rates for 61 specific drug-related problems occurring in three State Medicaid programs (Maryland, Iowa, and Washington) from 1989 through 1996 and a fourth (Georgia) from 1994 through 1996. The study is descriptive and exploratory. Although we discuss factors that might influence the DUR failure rates, no hypotheses are formally tested. Our DUR criteria are widely accepted as indicating significant preventable problems, but they are not exhaustive.

We hope to achieve two objectives in presenting these data. The first is to demonstrate the utility of a consistent DUR screening program in producing basic epidemiological data on drug-use problems in large populations. To our knowledge this is the first application of such a screener program for Medicaid drug-use data across States. It also represents the first time that longitudinal data on DUR failure rates have been published for these Medicaid programs. Our second objective is to help lay the groundwork for future analytic studies that can explain why State Medicaid programs exhibit highly divergent patterns in problematic prescribing. We believe that readers will find our results provocative and hope that this will generate interest in both further exploration of the problems and new tools to address them.

The genesis of this study was a provision in OBRA 1990 that mandated Medicaid demonstrations of innovative approaches to DUR. Between 1992 and 1996, two States implemented DUR demonstration programs under the HCFA mandate—paying pharmacists for cognitive services (Washington) and an on-line prospective

DUR (Iowa). Concurrently, a team of evaluators led by Abt Associates, Inc., designed an analysis plan that included recruitment of two non-demonstration States (Maryland and Georgia) to serve as comparator States (Abt Associates, Inc., 1989). One component of the evaluation compared drug policies, rates of drug use, and the frequency of drug-use problems in the four study States over the period 1989-96.

Of the four States, only Iowa and Maryland had retrospective DUR programs in place in 1989. Georgia and Washington both implemented such programs in 1993. Maryland was the first to implement a prospective DUR program (January 1993), followed by Iowa (June 1994, but only in demonstration pharmacies),¹ and Washington in March 1996. As of this writing, Georgia had not implemented a prospective DUR program.

Other regulatory policies that might be expected to affect drug use in the four States include formularies, prescription limits, and prior-authorization requirements. Georgia and Washington had closed formularies in 1989 and 1990 but dropped them as a result of the OBRA 1990 prohibitions. Maryland adopted an open formulary in 1991, followed by Washington (1996), Iowa (1997), and Georgia (1997). Of the four States, only Georgia places limits on the number of prescription medications that Medicaid recipients can fill in a month (five per month except by prior authorization). Both Iowa and Georgia have targeted prior-authorization programs. Beginning in 1990 Georgia has limited the maximum number of anxiolytic benzodiazepines to three prescriptions per calendar year without prior approval. Since 1992 Iowa has required prior authorization on the use of histamine-2 receptor antagonists (H2RAs), sucralfates, omepra-

zole, misoprostal, single-source non-steroidal anti-inflammatory drugs (NSAIDs), and benzodiazepines.

METHODS

The 61 DUR criteria screened for this study were originally developed by expert clinical panels convened by the University of Maryland and the Philadelphia College of Pharmacy and Science (UM/PCPS) under a cooperative agreement with HCFA (Knapp and Erwin, 1992). The UM/PCPS project was designed to produce a comprehensive and valid set of screening criteria appropriate for conducting outpatient DUR in State Medicaid programs. Since 1993 these criteria have undergone periodic reviews and updates by an expert panel of physicians and pharmacists associated with the DUR demonstration project previously described. In addition, each revision was further reviewed by a 10-member technical advisory group of recognized clinicians and researchers in pharmaceutical care to ensure that the DUR criteria continued to screen for significant problems backed by sound scientific evidence. The latest update of these criteria (January 1996) was used to produce the findings reported in this article. A detailed discussion of the UM/PCPS DUR criteria, the updating process, and the computer algorithms used to implement the screens is available in the Abt final report (Abt Associates, Inc., 1998).

The screening process involved a claim-by-claim review of all Medicaid prescriptions filled in the study States during the 8-year period (3 years in the case of Georgia) for the following drug categories: angiotensin-converting enzyme (ACE) inhibitors, antidepressants, antipsychotics, benzodiazepines, calcium channel blockers, digoxin, H2RAs, and NSAIDs. The drug problems screened include overdosing, excess duration of treatment, drug-

¹ The Iowa Medicaid program implemented a prospective DUR program statewide in July 1997.

drug interactions, duplicative therapy, contraindications, and initial therapy. A description of each screening criterion is shown in Table 1.

The screening process produced four types of outputs: (1) annual counts of all prescription claims for drug products reviewed by the screener; (2) annual counts of persons using these drugs; (3) annual counts of claims failing each of the 61 screening criteria; and (4) counts of persons with one or more screen failures per year by criterion.

These data were aggregated to produce annual DUR failure rates at both the claim and the person level by year and State. For Maryland, Iowa, and Washington, failure rates were computed for the entire 8-year period of 1989-96. Georgia presented a problem because the State did not require that pharmacists include the number of days supply on prescription bills until the middle of 1993. Because this number is a necessary element to run the screening program, Georgia DUR screening results are only reported for 1994 through 1996. The claim-level rates were computed by dividing the number of screen failures for each DUR criterion by the total number of claims screened for that criterion. Person-level screen failure rates were calculated by taking the ratio of the number of persons failing a specific criterion at least once in the year to the total number of persons using drugs in that criterion set in that same year.

The reason for producing two sets of rates is that each addresses different policy questions. Claims-based failure rates provide important information to DUR systems designers. Because all prospective DUR systems generate alerts at the claims level, this is a logical place to test system performance. DUR screens that generate “too many” alerts may be misrepresenting common therapeutic interventions as problematic when they are not (false positives).

Screens that generate few hits may be too tightly configured, missing the true extent of the underlying problem (poor sensitivity). Also, it is generally believed that pharmacists develop patterned responses to DUR screen alerts based on the frequency of the alert, the type of drug problem, and the source of the prescription (Armstrong and Denmark, 1998). These issues, too, are best examined at the claims level.

DUR screening at the person level gets to the heart of why on-line prospective DUR systems were developed—to alert pharmacists so that they can intervene and correct potential drug-related problems before they occur. If DUR programs are successful in this regard, the evidence should be observable at the person level. Because the risk of drug misadventures is not evenly distributed across the entire population, person-based screening should be targeted to known (or suspected) high-risk groups. For this project, DUR screening was performed for the entire Medicaid population in each State and for three high-risk subgroups—community-dwelling elderly,² adult diabetic persons, and multiple-drug users. The community-dwelling elderly subgroup was defined as Medicaid recipients age 65 or over with no nursing home claims during the year. Adult diabetic persons were defined as people 18 years of age or over with a diagnosis of diabetes (*International Classification of Diseases, 9th Revision* code 250.xx) and/or a prescription for an oral hyperglycemic agent. Multiple-drug users were defined as recipients filling prescriptions in three or more different therapeutic classes during the year.

The DUR screening results are provided in their entirety in the Abt report as 64 four-page tables, 16 for each State (Abt Associates, Inc., 1998). The number of per-

² The focus on community-dwelling rather than institutionalized elderly was driven by the fact that the HCFA DUR demonstrations did not include nursing home residents.

Table 1
Summary of UM/PCPS Drug Use Review Criteria, 1995 Update

Therapeutic Class	Dosage	Duplicative Therapy	Drug-Drug Interaction	Duration	Contraindication or Initial Therapy
ACE Inhibitors	Maximum daily dose per dosage table	Concurrent use of multiple ACE inhibitors	ACE inhibitor with potassium-sparing diuretics and/or potassium supplements; or with lithium; or with digoxin	—	—
Antidepressants	Maximum daily dosage per dosage table	Concurrent use of 2 or more antidepressants; or with stimulants	Antidepressant with amitriptyline, nortriptyline, imipramine, or clomipramine with another antimuscarinic or antispasmodic agent; or with 2 or more antimuscarinics or antispasmodics; or with MAO inhibitors; or tricyclics or maprotiline with guanethidine and/or guanadrel; or with clonidine; or imipramine and type 1 antiarrhythmics; or amoxapine and antiparkinsonians; or fluvoxamine with astemizole or terfenadine	—	—
Antipsychotics	Maximum adult daily dose per dosage table; maximum pediatric daily dose	Concurrent use of 2 or more antipsychotics	Antipsychotics with 3 or more antimuscarinic or antispasmodics; or with antiparkinsonian; or with levodopa	Maximum daily dose for more than 6 months without dose decrease	Not recommended for pediatrics
Benzodiazepines	Maximum adult daily dose per dosage table	Concurrent use of 2 or more benzodiazepines, except clonazepam; or with other anxiolytic drugs; or with sedative/hypnotics; or with antihistamines	—	Maximum daily dose for more than 30 days or lower dose for more than 4 months; estazolam, flurazepam, temazepam, and triazolam more than 35 days	Long half-life benzodiazepine with age over 65 years or in long-term care facility

See notes at end of table.

Table 1—Continued
Summary of UM/PCPS Drug Use Review Criteria, 1995 Update

Therapeutic Class	Dosage	Duplicative Therapy	Drug-Drug Interaction	Duration	Contraindication or Initial Therapy
Calcium Channel Blockers	Maximum adult daily dose per dosage table	Concurrent use of 2 or more calcium channel blockers	Calcium channel blockers with cyclosporine; verapamil with beta-adrenergic blocking agent; addition of digoxin to maintenance verapamil; addition of verapamil to maintenance digoxin; verapamil with disopyramide or flecainide	Nimodipine for more than 21 days	Bepridil approved for only second-line agent for angina
Digoxin	Maximum adult daily dose per dosage table	Concurrent use of 2 or more digitalis glycosides	Amiodarone/Quinidine added to givoxin without dosage drop of digoxin; high-dose digoxin added to Amiodarone, quinidine or propafenone; cyclosporine added to digoxin	—	—
H2RAs	Maximum adult daily maintenance dose per dosage table	Concurrent use of 2 or more H2RAs; or with misoprostol, omeprazole, or sucralfate; or with sucralfate on the same day	H2RA with ketoconazole; cimetidine with phenytoin, methylxanthines, warfarin, or procainamide	Maximum of 56 days for acute treatment; maximum of 93 days for GERD diagnosis	—
NSAIDs	Maximum adult daily dose per dosage table	Concurrent use of 2 or more NSAIDs	NSAID with high-dose methotrexate; select NSAIDs with oral anticoagulants; NSAIDs with lithium; phenylbutazone with oral antidiabetics; salicylates with oral antidiabetics; triamterene added to indomethacin; salicylates and probenecid; ketorolac with probenecid	Maximum of 7 days for mefenamic acid/phenylbutazone	Phenylbutazone is second-line therapy (removed from market in 1995)

NOTES: UM/PCPS is University of Maryland and Philadelphia College of Pharmacy and Science. ACE is angiotensin-converting enzyme. MAO is monoamine oxidase. H2RA is histamine-2 receptor antagonist. GERD is gastroesophageal reflux disease. NSAID is non-steroidal anti-inflammatory drug. SOURCE: Abt Associates, Inc., 1998.

sons reviewed per screening criterion, State, or year varies from 0 to more than 366,000. The number of claims reviewed ranges from 0 to 946,000. The screen failure rates encompass the entire spectrum from 0.00 to 100 percent. Typical cell sizes for the person-based statistics range between 10,000 and 50,000 recipients screened, with the numbers gradually increasing from 1989 until 1995, after which there are slight declines. Cell sizes in the prescription claims tables typically range between 50,000 and 200,000 claims reviewed per screening criterion per year.

To help make sense of the sheer volume of these data, the Abt report provided sections on major comparative patterns evident in drug utilization and screen failure rates: low versus high, rising versus falling, differences among the States, differences among risk groups, variation by drug class, and type of drug problem. All but the drug class sections are provided in this article. The analysis for all tables is purely descriptive. No statistical tests of significance are performed.³ No multivariate models are estimated. More attention is devoted to findings at the person level than at the claims level because that is where the potential health consequences are most evident.

DRUG USE BY THERAPEUTIC CLASS

Table 2 presents State-level summary data on drug use in each of the eight therapeutic categories reviewed for 1989 (1994 for Georgia) and 1996. As shown in this table, the number of Medicaid recipients with one or more prescription fills in the drug category for the year. Table 2 also shows the average annual number of fills for persons using

³ Strictly speaking, all differences in rates are “statistically significant” because the data represent population values rather than sample values. No attempt is made to generalize the results to other State Medicaid programs or to non-Medicaid populations.

products in the drug class. For the most part, recipient counts reflect both the relative sizes of the four State Medicaid programs (Iowa being the smallest and Georgia the largest) and the growth in Medicaid populations over time. More Medicaid recipients use NSAIDs than use drugs in any of the other classes. The drug class with the fewest users is digoxin (except in Georgia where there were fewer benzodiazepine users). The most dramatic growth in recipient counts from 1989 to 1996 occurred in the ACE inhibitor class. The number of ACE inhibitor users more than doubled in Iowa and Washington and tripled in Maryland during this period.⁴ In only one instance—digoxin users in Washington—were there fewer recipients in 1996 compared with 1989.

The average prescription counts per user by drug class shown in Table 2 indicate very different drug utilization patterns among the four States. Some of the variation is undoubtedly attributable to differences in the composition of the Medicaid populations, some to State Medicaid policies, and some to physician prescribing behaviors. Overall, utilization rates among users of these eight drug groups are highest in Iowa and lowest in Georgia. The difference between these two States is quite dramatic. The average number of prescriptions per person/year in Iowa during 1996 was higher in every class—50 percent higher in the case of antipsychotics, benzodiazepines, and NSAIDs. Not unexpectedly, antipsychotics exhibited the highest mean number of fills per year in any drug class. In Iowa and Washington, the average recipient of antipsychotic medications filled 11 claims in 1996 or nearly 1 per month. Although more Medicaid recipients use NSAIDs than any other class of drug, the average level of NSAID use is

⁴ During this time period, ACE inhibitors became first-line therapy for high blood pressure among diabetic persons and were also more commonly used for other hypertensive patients.

Table 2
Number of Recipients With Prescriptions and Average Prescription Claims per User, by Drug Class, State, and Year

Drug Category	Georgia		Iowa		Maryland		Washington	
	1994	1996	1989	1996	1989	1996	1989	1996
ACE Inhibitors	41,773	48,257	4,436	11,286	7,533	24,458	8,040	19,272
Antidepressants	51,023	51,818	12,947	19,684	10,955	22,244	25,511	33,492
Antipsychotics	40,940	42,965	11,453	14,513	13,160	23,855	18,875	21,932
Benzodiazepines	8,203	8,303	14,573	18,788	19,673	29,136	20,193	28,920
Calcium Channel Blockers	64,673	66,895	6,928	13,585	11,782	33,930	10,641	21,404
Digoxin	28,347	26,961	6,855	7,944	5,986	10,988	10,416	9,984
H2RAs	71,730	87,172	12,284	20,069	16,882	34,228	28,547	34,276
NSAIDs	231,872	207,807	30,953	43,430	39,455	63,645	62,187	58,201
ACE Inhibitors	6.4	6.8	6.7	7.7	Average Prescription Claims	5.7	5.5	7.2
Antidepressants	4.8	4.9	6.5	6.4		5.5	5.4	6.0
Antipsychotics	6.6	7.1	9.0	11.0		9.3	6.9	11.1
Benzodiazepines	4.2	4.5	6.9	6.9		6.2	5.0	5.7
Calcium Channel Blockers	6.8	7.2	7.0	7.9		6.0	5.7	7.3
Digoxin	6.9	7.1	8.1	8.4		6.2	4.1	6.5
H2RAs	2.8	4.5	5.0	5.3		4.7	4.4	5.0
NSAIDs	2.4	2.4	3.7	3.1		2.7	3.1	3.3

NOTES: ACE is angiotensin-converting enzyme. H2RA is histamine-2 receptor antagonist. NSAID is non-steroidal anti-inflammatory drug.

SOURCE: Medicaid drug data from Georgia, Iowa, Maryland, and Washington, 1989, 1994, and 1996.

low—between two and three prescription fills per person and year across the States.

The table shows a definite rising trend in drug usage. The greatest gains were recorded in the antipsychotic, ACE inhibitor, and calcium channel blocker categories. Washington far outpaced the other States in terms of rising numbers of claims per user. Between 1989 and 1996, annual usage rates for antipsychotics rose from 6.9 to 11.1 prescriptions, a 61-percent increase. Digoxin use increased by 59 percent, ACE inhibitors by 31 percent, and calcium channel blockers by 28 percent. There were a few instances of declining average utilization rates during the 8-year period: antidepressants and NSAIDs in Iowa and Maryland and digoxins in Maryland.

DUR SCREEN FAILURE RESULTS

Consistently Low Failure Rates

Low DUR failure rates are to be applauded because they indicate either that practice conforms to expert opinion or that the problems themselves are rare.⁵ What constitutes low prevalence in screen failure rates? For 3 of the 61 criteria screened, there were no instances of DUR failures in any State, year, or population subgroup. These three are all drug-drug interactions rated as having class 1 significance in Hansten's Interactions and Updates: antidepressants with a monoamine oxidase (MAO) inhibitor, amoxapine with an antiparkinsonian agent, and ketorolac with probenecid. Only in the latter instance were the number of claims or persons screened low in absolute numbers. The number of Medicaid recipients using either ketorolac or probenecid ranged from a low of 373 in Iowa (1989) to a high of 1,727 in

⁵ This assumes, of course, that the DUR criteria are valid and have high specificity. The specificity of all PSU screening criteria is high (Stuart, Ahern, and Coulson, 1994).

Washington (1991). Even rarer is use of nimodipine. Screening for nimodipine duration of therapy identified just a handful of users (20 recipients in Georgia in 1996) and even fewer instances of DUR failure over the entire panel of States and years. By contrast, between 12,000 recipients (Maryland, 1989) and 80,000 recipients (Georgia, 1996) were screened for the antidepressant-with-MAO-inhibitor interaction. Finding none in such a large population base certainly confirms that practitioners are aware of the problem and abide by the specified drug-use criteria.

Raising the threshold of "low" to less than 1 screen failure in 400 claims or persons reviewed (less than 0.25 percent failure rate) in any State or year resulted in 13 of the 61 DUR criteria qualifying for low-failure-rate status (Table 3). All but 2 of these 13 criteria (use of bepridil or phenylbutazone as first-line therapy) reference drug-drug interactions, and all can produce severe consequences. These drug problems cluster in three drug groups: antidepressants, NSAIDs, and digoxins (by contrast, none of the DUR criteria in the ACE inhibitor, benzodiazepine, or H2RA drug groups met this standard for low prevalence of screen failures). As one would expect, there is a close match between failure rates at the claim and person levels. In every case but one (the interaction of phenylbutazone with oral antidiabetics), low prevalence of claims failures is associated with low prevalence in failure rates at the person level. And even for that criterion, the highest recorded person-level prevalence was 0.39 percent (Iowa, 1989).

The ketorolac-with-probenecid interaction remains the only low-failure-rate criterion in which the number of persons and claims screened was itself small. At least 4,000 recipients and 30,000 claims were screened in each State and year for the other 12 criteria (the typical range was

Table 3
Rates for DUR Criteria With Consistently Low Failure Rates: Four Study States, 1989-96

DUR Criterion	Range of Failure Rates per Claim	Range of Failure Rates per Person
Antidepressants		
		Percent
Antidepressants With MAO Inhibitor	0.00-0.00	0.00-0.00
Amoxapine With Antiparkinsonian	0.00-0.00	0.00-0.00
Tricyclic Antidepressant or Maprotiline With Guanethidine/Guanadrel	0.00-0.11	0.01-0.07
Antipsychotics		
Antipsychotic With More Than 3 Antimuscarinics or Antispasmodics	0.00-0.06	0.00-0.04
Calcium Channel Blockers		
Add Verapamil to Digoxin	0.01-0.09	0.00-0.16
Bepridil as Initial Therapy	0.00-0.00	0.00-0.03
Digoxins		
Cyclosporine Added to Digoxin	0.00-0.02	0.00-0.09
Digoxin Plus Amiodarone/Quinidine Without Dosage Drop	0.00-0.05	0.01-0.10
Multiple Digoxins	0.00-0.09	0.00-0.04
NSAIDs		
Ketorolac With Probenecid	0.00-0.00	0.00-0.00
NSAID With High-Dose Methotrexate	0.00-0.01	0.00-0.01
Phenylbutazone as Initial Therapy	0.00-0.07	0.00-0.25
Phenylbutazone With Oral Antidiabetic	0.00-0.13	—

NOTES: DUR is drug utilization review. MAO is monoamine oxidase. NSAID is non-steroidal anti-inflammatory drug. "Consistently low failure rates" indicate DUR criteria for which less than 0.25 percent of claims or persons reviewed failed the criterion in question in every State and year considered.

SOURCE: Medicaid drug data from Georgia, Iowa, Maryland, Washington, 1989-96.

between 10,000 and 40,000 recipients and between 70,000 and 400,000 claims screened). Here again, with such a large risk exposure, the fact that there were so few DUR failures is very encouraging. One reason for this may be that certain criteria, such as cyclosporine added to digoxin or NSAID with high-dose methotrexate, are associated with drugs generally prescribed by specialists, which implies controlled use. Other criteria, such as those with digoxin, represent a very narrow therapeutic index that generally encourages careful prescribing by all clinicians.

Consistently High Failure Rates

At the other end of the spectrum are DUR criteria that are frequently violated. Experts may disagree about what constitutes a "high" DUR failure rate, but few would take issue with the 10 criteria shown in Table 4. To be selected for this exclusive list, the failure rate had to exceed 8.0

percent of all claims or persons screened in every State and year of the study. Even with this stringent standard for inclusion, the lower bound failure rates ranged from 8.1 percent (the concurrent use of an H2RA with misoprostol, omeprazole, or sucralfate) to 53.4 percent (excessive duration of therapy for estazolam, flurazepam, temazepam, or triazolam).⁶ At the upper bound, failure rates approached 70 percent (excessive duration of therapy with mefenamic acid or phenylbutazone).

High failure rates do not necessarily mean that large numbers of persons failed the screening criteria. For example, fewer than 150 individuals failed the mefenamic acid or phenylbutazone duration screen in any State or year. Fewer than 300 failed the estazolam, flurazepam, temazepam, or triazolam duration screen in any of the 3 years

⁶ The duration-of-therapy screens for benzodiazepines are complex. The duration of therapy for estazolam, flurazepam, temazepam, or triazolam is set at 35 days, whereas the general benzodiazepine duration-of-therapy screen for all drugs in this class (including the four just named) is dose-related. It is thus possible that a claim may fire both screens simultaneously.

Table 4
Rates for DUR Criteria With Consistently High Failure Rates: Four Study States, 1989-96

DUR Criterion	Range of Failure Rates per Claim	Range of Failure Rates per Person
Antipsychotics		
Multiple Antipsychotics	—	Percent 9.9-16.3
Antidepressants		
Antidepressant Dosage	11.8-21.5	14.3-26.6
Benzodiazepines		
Benzodiazepine With Anxiolytic	8.8-16.0	—
Benzodiazepine Duration for Estazolam, Flurazepam, Temazepam, Triazolam	27.4-48.6 53.4-66.4	26.0-41.4 15.6-56.0
Long Half-Life Benzodiazepine for Person Age Over 65	21.8-39.9	20.6-37.2
H2RAs		
H2RA Adult Dosage	11.9-35.5	9.8-39.6
H2RA with Misoprostol, Omeprazole, Sucralfate	—	8.1-14.7
NSAIDs		
Mefenamic Acid or Phenylbutazone Duration	29.4-70.0	30.0-68.3
NSAID Dosage	—	11.2-22.2

NOTES: DUR is drug utilization review. H2RA is histamine-2 receptor antagonist. NSAID is non-steroidal anti-inflammatory drug. "Consistently high failure rates" indicate DUR criteria for which more than 8.0 percent of claims or persons fail the listed criterion in every State and year considered.

SOURCE: Medicaid drug data from Georgia, Iowa, Maryland, Washington, 1989-96.

of data available from Georgia. On the other hand, between 7,000 and 15,000 Medicaid recipients in Iowa, Washington, and Maryland failed the benzodiazepine duration-of-therapy criterion each year of the study. Large numbers of individuals also failed the NSAID and H2RA dosage screens. In terms of risk exposure, the number of recipients screened for these high-failure-rate criteria ranged from 10 in Iowa for the mefenamic acid/phenylbutazone duration screen (1995) to 250,747 in Georgia for NSAID adult dosage (1995).

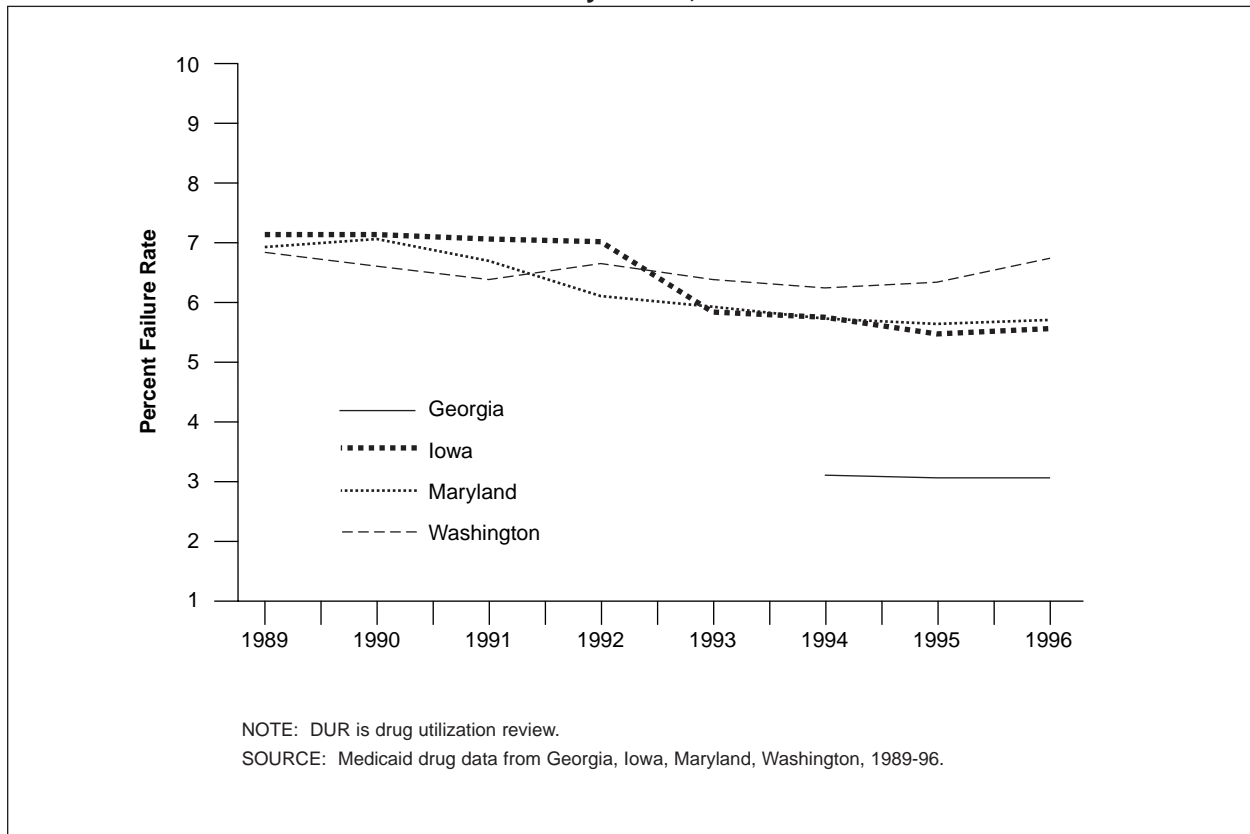
What might account for such extreme values? Four of the 10 criteria associated with high failure rates relate to benzodiazepine use. Seven are associated with high-dosage or excessive-duration-of-therapy criteria. The benzodiazepine drug class presents an interesting case because these drugs have been subject to intense scrutiny in the scientific literature, especially concerning misuse of these medications among the elderly. Duration and dosage are types of criteria particularly prone to either subjective clinical

judgment or high false-positive rates. Physicians may feel more comfortable in intentionally prescribing outside usual dosage or duration parameters than they would, say, in prescribing drugs with known, serious interactions. It is also possible that some of the reported DUR problems in Table 4 are inflated: Computer interpretations of clinical criteria are sometimes insensitive to common prescribing and drug-use behaviors (e.g., early refills are flagged as high dosage, or a treatment change from monotherapy to polytherapy is flagged as duplicative therapy). Whether the problems are real or artifactual, they clearly should be addressed if DUR screening is to be a useful quality-improvement tool.

Summary Measures of DUR Failure Rates

Figures 1-4 show trends in DUR screen failures averaged across all 61 screening criteria. The average is computed by dividing the sum of recipients screened for all

Figure 1
Average Failure Rate for 61 DUR Screening Criteria for All Medicaid Recipients:
Four Study States, 1989-96



criteria by the sum of screen failures in each State and year.⁷ Very clear patterns emerge from these graphs.

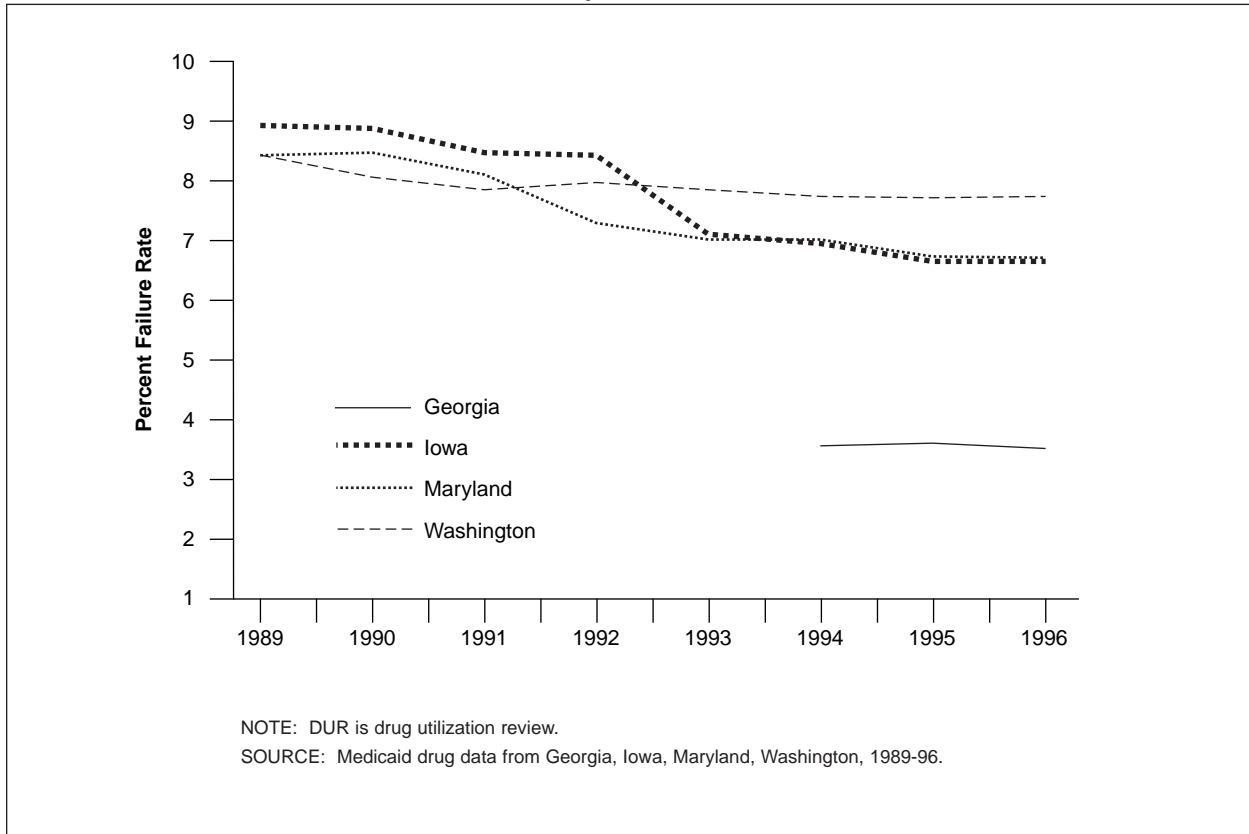
Average DUR failure rates for the entire Medicaid population in each State are portrayed in Figure 1. The Iowa, Maryland, and Washington rates cluster between 5.5 percent and 7.2 percent of persons with one or more screen failures per year. Georgia's failure rates are half that, averaging just 3.1 percent per year. Both Iowa and Maryland exhibited significant improvements in DUR failure rates during

the 8 years of the study. In Iowa the rate fell from 7.2 percent to 5.5 percent, with most of the improvement registered in 1993, 1 year after the State adopted its prior-authorization program for clozapine, H2RAs, sucralfate, omeprazole, misoprostol, single-source NSAIDs, and benzodiazepines. The improvement in Maryland was more gradual, with mean failure rates declining from 7.0 percent in 1989 to 5.7 percent in 1996. Washington ended 1996 about where it started in 1989, with DUR failure rates averaging 6.8 percent per year.

The trends in DUR failures for multiple-drug users (Figure 2) and adult diabetic persons (Figure 3) map closely those for the Medicaid population as a whole, except that the rates are higher across the board by 1 to 2 percentage points. These results are not unexpected. Persons using drugs

⁷ Computing the average in this way weights each screening criterion by the number of persons affected. An alternative method of computing an all-screen average is to sum the failure rates for each screening criterion and then divide the result by 61. This latter method gives each criterion equal weight regardless of the number of persons affected. The empirical results are similar either way. The equal-weight method produces higher failure rates by 1 to 2 percentage points, but the relative rankings of States and years are unaffected by choice of method.

Figure 2
Average Failure Rate for 61 DUR Screening Criteria for Multiple-Drug Users:
Four Study States, 1989-96

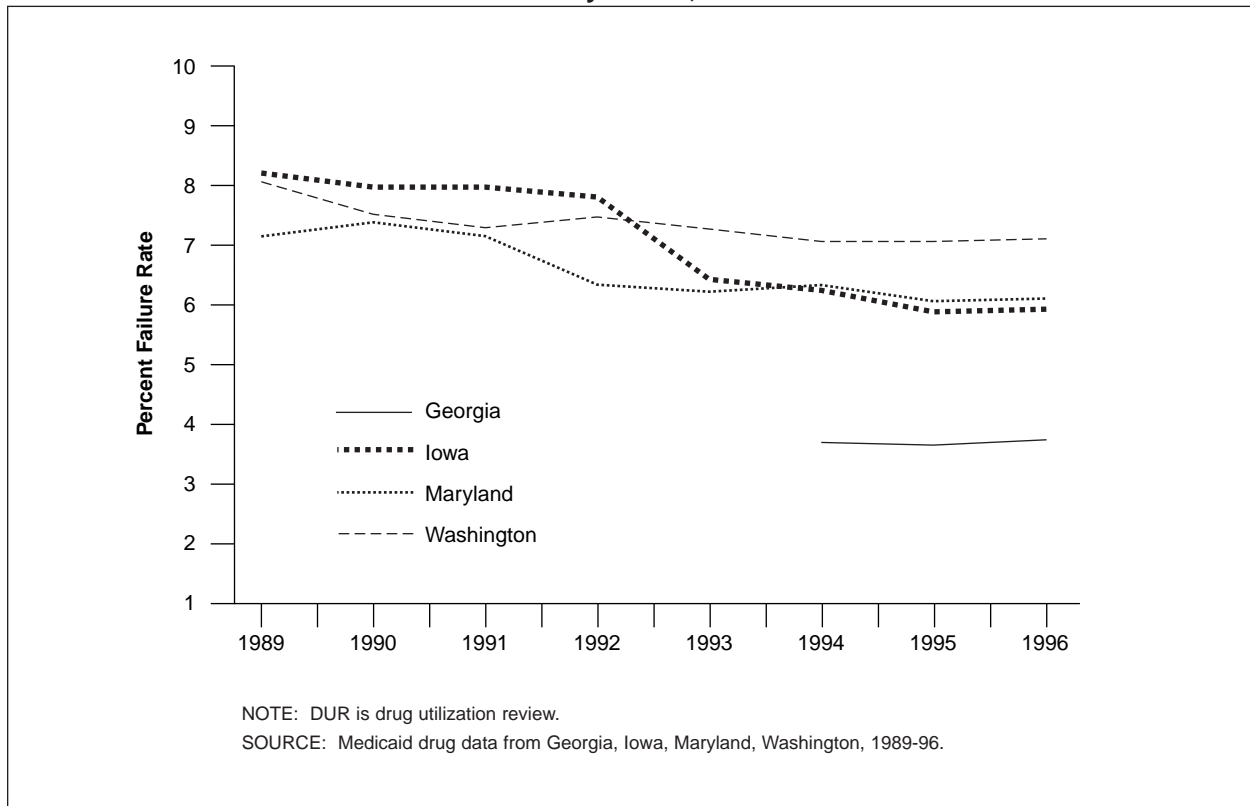


within several therapeutic categories are at increased risk for various DUR criteria violations, particularly drug-drug interactions. Although only 1 DUR screen among the 61 specifically targets diabetes medications (phenylbutazone with oral antidiabetic), the complications of this disease mean that adult diabetic persons typically take various medications that also increase their risk for DUR failures.

Figure 4 displays summary data on screen failure rates for the community-dwelling elderly subpopulation in each State. The findings shown here are perhaps the most dramatic of any presented thus far. Here too, Georgia led all States with the elderly failing DUR screens at an average rate of 2.2-2.4 percent annually (slightly higher than for the entire Medicaid population). The real story is in

the other three States, where DUR failure rates for the elderly fell precipitously during the 8-year time span. In Iowa the failure rate for elderly recipients in 1989 was equivalent to the Medicaid population as a whole (7.1 versus 7.2 percent, respectively). But by 1996 the mean failure rate for the elderly was only 4.5 percent, a full percentage point lower than the population average. In Maryland and Washington, 1989 failure rates for the elderly were much higher than the population averages (8.0 versus 7.0 percent in Maryland, 8.0 versus 6.9 percent in Washington). By 1996, however, the rates for the elderly were below the population levels (5.5 versus 5.7 percent in Maryland, 6.2 versus 6.7 percent in Washington). This is a considerable achievement. Although there are probably many contributing factors involved here,

Figure 3
Average Failure Rate for 61 DUR Screening Criteria for Adult Diabetic Persons:
Four Study States, 1989-96



the considerable attention devoted to drug problems and aging in both the scientific and lay press since the late 1980s must surely have played a major role.

SUMMARY AND CONCLUSIONS

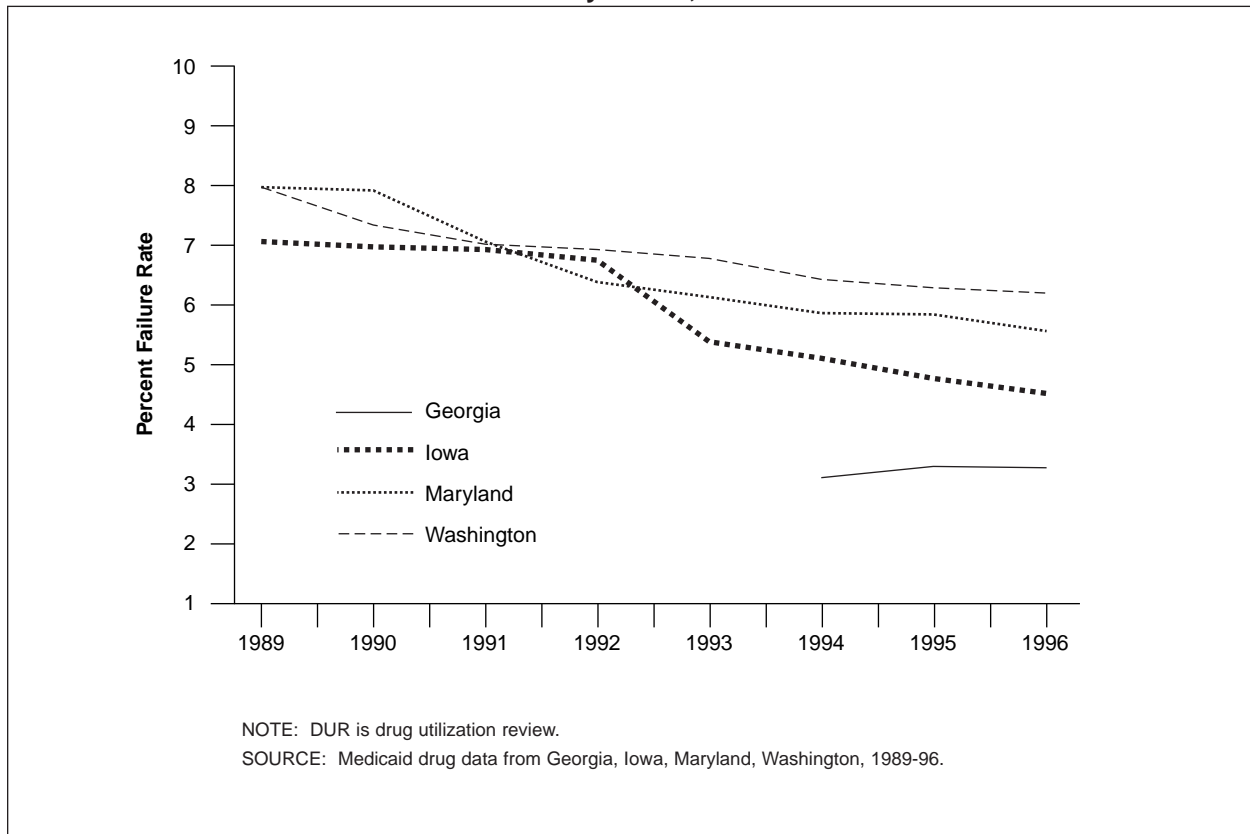
In this article we have presented descriptive findings from screening 8 years of drug claims for Medicaid programs in Maryland, Iowa, and Washington and 3 years in the case of Georgia, using a consistent DUR screener system developed by the Pennsylvania State University. The findings represent a rich source of data for policy analysis of drug-related problems in these programs. Following is a summary of the most significant findings.

- Georgia had the lowest and Iowa the highest rates of drug utilization rates within the eight categories reviewed. In

some drug classes, usage rates were 50 percent higher in Iowa than Georgia.

- Drug use rose in all States during the study period. The largest increases were recorded for antipsychotics, ACE inhibitors, and calcium channel blockers.
- Georgia had the lowest mean failure rate, at 3.1 percent of users in 1996, compared with rates of 5.4 to 6.7 percent in the other States that year.
- Significant improvement in DUR failure rates was recorded in Iowa (from a mean of 7.2 percent of persons failing DUR screens in 1989 to 5.5 percent in 1996) and Maryland (from a failure rate of 7.0 percent to 5.7 percent).
- Much of the improved performance is attributable to declining DUR failures among community-dwelling elderly. During the 8 years of the study, average DUR failure rates in this population

Figure 4
Average Failure Rate for 61 DUR Screening Criteria for Community-Dwelling Elderly:
Four Study States, 1989-96



declined by 35 percent in Iowa, 31 percent in Maryland, and 23 percent in Washington.

- Multiple-drug users and adult diabetic persons had DUR failure rates 1 to 2 percentage points higher than the Medicaid population as a whole. Both groups showed improvement during the study period but not as dramatically as for the community-dwelling elderly.
- DUR failure rates vary significantly by drug group and DUR criteria. Ten of the 61 criteria were associated with consistently high failure rates exceeding 8.0 percent of persons and claims in every State and year reviewed. Thirteen were associated with consistently low failure rates below 0.25 percent of persons or claims per year.

- The drug groups with the highest failure rates cluster in the benzodiazepine, H2RA, and NSAID classes. DUR criteria relating to dosage and duration-of-therapy problems are also associated with high failure rates.
- In only one instance (1993 in Iowa) was there any sharp decline in DUR failure rates in the year following adoption of a State DUR policy. This does not mean that State DUR policies are ineffective but rather that the impact of these policies may be gradual and difficult to detect in descriptive time-series data. The differentials in DUR failure rates highlighted here raise important policy questions. Why did failure rates fall substantially in Iowa and Maryland but not in Washington (except for the elderly)? What

specific factors explain the dramatic reductions in DUR failures among community-dwelling elderly? Might not these same factors be brought to bear to improve the drug-use profiles of other high-risk groups? Why are the experts' opinions regarding appropriate dose and duration criteria followed much less often than their judgments regarding drug interaction and duplicative-therapy criteria? Perhaps selected screening parameters should be empirically derived to better reflect actual prescribing practices. But if so, which ones?

Data alone cannot answer these questions. Descriptive studies such as this one are better at generating hypotheses than at testing them. It remains for future work to design research protocols in conjunction with DUR databases to take full advantage of this valuable information source.

REFERENCES

Abt Associates, Inc.: *Evaluation of Drug Use Review Demonstration Projects*. Contract Number 500-93-0002. Final Report. Cambridge, MA. October 28, 1998.

Alliance for Aging Research: *When Medicine Hurts Instead of Helps: Preventing Medication Problems in Older Persons*. Washington DC. 1998.

Armstrong, E., and Denemark, C.: How Pharmacists Respond to On-Line, Real-Time Alerts. *Journal of the American Pharmacy Association* 38:149-154, March-April 1998.

Bates, D.W., Spell, N., and Cullen, D.J.: The Costs of Adverse Drug Events in Hospitalized Patients. *Journal of the American Medical Association* 277(4):307-311, January 22-29, 1997.

Bootman, J.L., Harrison, D.L., and Cox, E.: The Health Care Cost of Drug-Related Morbidity and Mortality in Nursing Facilities. *Archives of Internal Medicine* 157(12):2089-2096, October 13, 1997.

Johnson, J.A., and Bootman, J.L.: Drug-Related Morbidity and Mortality: A Cost-of-Illness Model. *Archives of Internal Medicine* 155(18):1949-1956, October 9, 1995.

Knapp, D., and Erwin, G.: *Screening Criteria for Outpatient Drug Use Review*. Final Report to HCFA. Baltimore, MD. University of Maryland School of Pharmacy, 1992.

Lazarou, J., Pomeranz, B.H., and Corey, P.N.: Incidence of Adverse Drug Reactions in Hospitalized Patients: A Meta-analysis of Prospective Studies. *Journal of the American Medical Association* 279(15):1200-1205, April 15, 1998.

Stuart, B., Ahern, F., and Coulson, N.E.: *Impact of Prescription Drug Coverage on Aged Pennsylvania Medicare Beneficiaries*. Final Report to HCFA under HCFA Cooperative Agreement Number 17-C-99392. 1994.

Reprint Requests: Bruce Stuart, Ph.D., University of Maryland School of Pharmacy, 100 North Greene Street, 6th Floor, Baltimore, MD 21201-1563. E-mail: bstuart@rx.umaryland.edu