A Randomized Outpatient Trial of a Decision-Support Information Technology Tool

Michael Apkon, MD, PhD, MBA; Jennifer A. Mattera, MPH; Zhengiu Lin, PhD; Jeph Herrin, PhD; Elizabeth H. Bradley, PhD; Michael Carbone, BA; Eric S. Holmboe, MD; Cary P. Gross, MD; Jared G. Selter, MD; Amy S. Rich, MPH; Harlan M. Krumholz, MD, SM

Background: Decision-support information technology is often adopted to improve clinical decision making, but it is rarely rigorously evaluated. Congress mandated the evaluation of Problem-Knowledge Couplers (PKC Corp, Burlington, Vt), a decision-support tool proposed for the Department of Defense’s new health information network.

Methods: This was a patient-level randomized trial conducted at 2 military practices. A total of 936 patients were allocated to the intervention group and 966 to usual care. Couplers were applied before routine ambulatory clinic visits. The primary outcome was quality of care, which was assessed based on the total percentage of any of 24 health care quality process measures (opportunities to provide evidence-based care) that were fulfilled. Secondary outcomes included medical resources consumed within 60 days of enrollment and patient and provider satisfaction.

Results: There were 4639 health care opportunities (2374 in the Coupler group and 2265 in the usual-care group), with no difference in the proportion of opportunities fulfilled (33.9% vs 30.7%; P = .12). Although there was a modest improvement in performance on screening/preventive measures, it was offset by poorer performance on some measures of acute care. Coupler patients used more laboratory and pharmacy resources than usual-care patients (logarithmic mean difference, $71). No difference in patient satisfaction was observed between groups, and provider satisfaction was mixed.

Conclusion: This study provides no strong evidence to support the utility of this decision-support tool, but it demonstrates the value of rigorous evaluation of decision-support information technology.

Arch Intern Med. 2005;165:2388-2394

MEDICAL DECISION MAKING integrates patient-specific data with medical knowledge under conditions of uncertainty. Computerized decision-support information technology (DSIT) tools are programs that help caregivers close gaps between knowledge and performance1,2 using various approaches, including alerts, feedback, interpretation, prognostic tools, and diagnostic aids. Although these tools are proliferating,3 few studies4-6 have rigorously evaluated their impact on ambulatory care. The DSIT tools often escape the scrutiny applied to drugs or devices,7,8 yet they can be costly and can cause both harm9,10 and benefit.3,11

See also pages 2329 and 2395

The Department of Defense is incorporating the DSIT tool Problem-Knowledge Couplers (PKC Corp, Burlington, Vt) into the second generation of its computerized medical record.12 To date, the Department of Defense has spent more than $15 million licensing this tool for use in pilot programs across the Military Health System,13 which provides health care to 8.9 million people in 75 military hospitals and more than 450 clinics. Couplers uses structured questions based on the patient’s chief complaint to elicit information from the patient and the provider. That information is linked to a proprietary database of medical knowledge that generates suggestions for appropriate patient care strategies, including options relevant to the patient’s diagnosis and treatment.14,15 To satisfy a congressional mandate to evaluate Couplers before widespread implementation, we conducted a randomized controlled trial (RCT) of the effect of this DSIT tool in ambulatory practice, examining its effect on quality, resource consumption, and patient and provider satisfaction.
Knox, Ky, and Mayport Branch Health Clinic, Mayport, Fla). We also evaluated historical control groups at each facility and a contemporaneous control group from a clinic not participating in the RCT. The facilities were chosen by the Department of Defense because of their strong leadership support for Couplers and ambulatory practices consistent with the Military Health System optimization model specifying 2 examination rooms and 3.3 support staff per primary caregiver.

PARTICIPANTS

Patients aged 18 years and older were eligible if they had scheduled appointments, could speak and read English, had not participated in Coupler sessions, were not scheduled for obstetric care, and had no emergency medical conditions. The study was approved by the institutional review boards of Walter Reed Army Medical Center, Washington, DC; National Naval Medical Center, Bethesda, Md; and Yale University. Informed consent was obtained from all participating patients and providers.

INTERVENTIONS

Couplers are available for a wide range of preventive health care needs and patient complaints, including common conditions. Patients randomized to use Couplers completed the one appropriate for their specific complaint or, when no condition-specific Coupler was appropriate, a generic History and Screening Coupler, replicating the Military Health System’s intended use for Couplers. Patients entered their medical histories into the Coupler tool with the assistance of a coordinator who was not associated with the study; 30 minutes was allocated for that process. Providers treating Coupler patients could enter additional information before reviewing Coupler outputs outlining diagnosis or treatment options. Patients in the usual-care group had no exposure to Couplers.

OUTCOMES

The primary study outcome, quality of care, was assessed using 24 health care process measures grouped into 2 categories: screening/prevention and acute/chronic disease management. Measures (“opportunities for quality care”) were selected before initiating the study and were drawn from recommendations of quality-focused national organizations, such as the Agency for Healthcare Research and Quality and the U.S. Preventive Services Task Force. For each patient, the processes (or opportunities) indicated at the index visit were tabulated based on the patient’s medical characteristics. Structured medical record abstraction and data from the Military Health System’s electronic medical system (Composite Healthcare System) were used to determine whether each of these opportunities was satisfied (“fulfilled”) within 60 days of the index visit. The primary outcome was the overall fulfillment opportunity rates of 0.3. We calculated needing 1680 patients (840 per group).

EXTERNAL CONTROL AND HISTORICAL CONTROL GROUPS

To determine whether providers’ exposure to Couplers may have affected the care of usual-care group patients, end points were also evaluated in a separate but similar clinic at one of the sites (the “external control” clinic). In the external control clinic, providers were not exposed to Couplers, although they knew that their patients may have been included in the quality measurement study. We also assessed all 3 clinics’ performance before the study by examining the medical records of a sample of historical patients at each (“historical control” groups) to account for baseline differences in quality between the study and external control clinics. Patients in the external control and historical control groups were randomly selected based on appointment times.

STATISTICAL ANALYSIS

We tested the primary hypothesis that fulfillment rates differed between study groups by comparing the likelihood of fulfillment using a Mantel-Haenszel χ2 test of homogeneity, stratified by physician and adjusted for clustering by patient. Second, we evaluated differences in fulfillment rates between study groups for screening/prevention opportunities combined, for acute/chronic opportunities combined, and for individual opportunities. To account for clustering of outcomes within patients, we used cluster-adjusted χ2 tests but stratified by clinic given the inadequate number of individual opportunity types per physician to stratify by physician. To account for differences in patient characteristics and to improve precision of estimates of effects, we tested the effectiveness of Couplers using a 3-level hierarchical logistic regression model incorporating opportunity type and patient and provider characteristics, including group assignment (Coupler or usual care) as a patient characteristic.

We evaluated differences in median resource use in each category and overall using the Wilcoxon rank sum test of equality of distribution. We developed a hierarchical linear regres-
sion model with the logarithm of total resource use as the outcome to account for a skewed resource distribution. We examined differences in patient satisfaction between groups using summary scales for 4 domains (during visit, care provider, personal issues, and overall assessment). We calculated reliability estimates (Cronbach’s α) for each domain scale and compared the scales across study groups using Wilcoxon rank sum tests of equality of distribution. Physicians were also surveyed about the tool. All statistical calculations were performed using Stata 8 (StataCorp, College Station, Tex) and HLM 5 (Scientific Software International Inc, Lincolnwood, Ill).

STUDY SAMPLE

A total of 2769 patients were screened at the 2 sites between April 22, 2004, and December 31, 2002 (Figure). Of these patients, 611 (22.1%) did not meet the eligibility criteria, 252 (9.1%) refused participation, and 4 (0.1%) were removed from the study at their request, resulting in a total of 1902 patients (88.1% of 2158 eligible patients) enrolled and randomized: 936 (49.2%) in the Coupler group and 966 (50.8%) in the usual-care group. Site 1 enrolled 998 patients (480 in the Coupler group and 518 in the usual-care group), and site 2 enrolled 904 patients (456 in the Coupler group and 448 in the usual-care group). Not all patients had a valid opportunity type for inclusion in the primary analysis; those with at least 1 valid opportunity (721 in the Coupler group and 704 in the usual-care group) differed from those with none: they were younger (P<.001), were more likely to be female (P<.001), and more often had acute or routine visit types (P<.001). Patients were analyzed on an intention-to-treat basis. There were no reported adverse events associated with this study.

The characteristics of the 1902 patients by randomization status and site are given in Table 1. The mean age, sex, and visit type were similar between the Cou-
pler and usual-care groups. There were slightly more active-duty status patients in the usual-care group (44.0% vs 38.6%). The proportion of missing or incomplete medical records was similar in both groups (16%) and was largely due to patients moving away from study sites or maintaining possession of their medical records. The mean number of opportunities per patient, the proportion of patients with at least 1 screening/prevention opportunity, and the proportion with at least 1 acute/chronic opportunity were similar in the 2 groups.

**OUTCOMES**

**Primary Analysis**

There were 4639 health care opportunities (2374 in the Coupler group and 2265 in the usual-care group) among study patients at their index visit. There was no significant difference in the proportion of health care opportunities fulfilled in the Coupler and usual-care groups. There were slightly more acute/chronic opportunities fulfilled in the Coupler group vs the usual-care group (34.8% vs 30.4%; \( P = .03 \)), with dietary and exercise counseling being primarily responsible for the difference (Table 2). In contrast to screening/prevention opportunities, there were slightly fewer acute/chronic opportunities fulfilled in the Coupler group vs the usual-care group (27.7% vs 32.6%), although this difference was not significant. In the analysis by opportunity, 5 opportunities showed significant differences (Table 2).

In the analysis by opportunity subgroup (screening/prevention or acute/chronic), there were slightly more screening/prevention opportunities fulfilled in the Coupler group vs the usual-care group (34.8% vs 30.4%; \( P = .03 \)), with dietary and exercise counseling being primarily responsible for the difference (Table 2). In contrast to screening/prevention opportunities, there were slightly fewer acute/chronic opportunities fulfilled in the Coupler group vs the usual-care group (27.7% vs 32.6%), although this difference was not significant. In the analysis by opportunity, 5 opportunities showed significant differences (Table 2).

**Supplemental Analyses**

Comparisons between a sample of historical patients at the site 2 study clinic and historical patients at the site 2 external control clinic found no differences in patient characteristics, number of opportunities, or opportunity fulfillment. However, the proportion of total opportunities fulfilled in the period before the study was much lower for patients in the RCT and external control clinics (12.2% and 16.5%, respectively) compared with the study period RCT usual-care group (42.9%) and...
Queries of the Composite Healthcare System database identified data for 92.5% of the study patients, with equal proportions of patients missing data across groups. No significant differences in patient satisfaction survey results in any domains of satisfaction as determined using the Wilcoxon rank sum test or multivariable models (Table 5).

We surveyed each provider in the study (n=12 [8 physicians, 3 physician assistants, and 1 nurse practitioner; 8 at site 1 and 4 at site 2]) to assess his or her satisfaction with Couplers. The strongest level of perceived satisfaction related to information quality: 75% agreed or strongly agreed that Couplers provides high-quality information. The strongest level of dissatisfaction related to time use, with 83% disagreeing or strongly disagreeing that Coupler use involves acceptable amounts of time. More than half of the providers also disagreed with the statements of benefits for medical decision making (70%), improved provider-patient interactions (61%), and overall benefits to patients (70%).

In this study, implementing a specific DSIT tool in a primary care setting did not substantially improve quality of care, decrease resource consumption, or improve satisfaction of patients or providers. In secondary analyses, the modest improvement in screening/prevention opportunities was offset by lower rates of acute/chronic opportunities. The results indicate the importance of a thorough evaluation of DSIT tools before assuming that they are effective based on face validity.

The relevance of any DSIT tool evaluation depends on the intended use of the tool. We designed this study in coordination with the developers of the tool to ensure that we were evaluating meaningful outcomes. Measurement across these 24 specific process measures allowed us to examine quality in a setting of diverse patients, diagnoses, and activities typical of ambulatory medical practice. Our method of aggregating unweighted measurements was recently used by McGlynn et al21 to measure the effectiveness of health care delivery in large populations.

Few studies of DSIT tools exist, yet the use of these tools proliferates. With the expansion of electronic health records, the integration of such tools into practice will likely accelerate. Past studies have often focused on validating tools’ suggestions by comparing their perfor-
mance with that of experts using case simulations22 or by examining the impact on the diagnostic accuracy of users.23 Several studies have shown that DSIT tools may improve a physician’s diagnostic performance,24,25 depending on characteristics of the case, the provider, and other factors.26 However, most studies focused on limited numbers of diagnoses or problems and evaluated a tool’s efficacy rather than its effectiveness influenced by human factors or its impact on other dimensions of care.

This study has several issues to consider. Providers in the study cared for both intervention and control patients. However, using a historical control cohort and a concurrent control clinic, we found no evidence that the lack of effect was due to an effect on the care of the control patients. There was a significant improvement in care for the external control and study clinics at site 2. This could reflect the knowledge that performance was being measured, or it could reflect some other change in clinic operations. Improvement of this magnitude did not occur at site 1. Without the historical control, it would not have been possible to exclude Coupler use in some patients, affecting the care of all patients. Without the external controls, the improvement compared with the historical controls could be erroneously used to corroborate that possibility. A related potential limitation of using only 2 study sites is reflected in the different underlying percentages of fulfilled opportunities. However, the effect of the intervention was consistent across sites. Although we documented a limited benefit of Couplers, we cannot exclude the possibility that it may have facilitated the diagnosis of rare conditions or enhanced the performance of less experienced physicians or nonphysicians in clinical roles. Finally, this study was conducted at military facilities, so the findings may not be generalizable. Nevertheless, we believe that this setting, with the strong administrative support and optimized clinical environment, would have been ideal to show the benefit of the intervention.

This study serves as a paradigm for measuring quality of care in a multidimensional manner across a diverse ambulatory practice and for evaluating the impact of a complex intervention, such as DSIT tool deployment, on that quality. Such a multidimensional evaluation is critical given that direct effects on clinical quality and process reliability are but one measure of the benefits and costs of interventions such as DSIT tool implementation, new protocol introduction, new educational initiatives, or workflow redesign efforts. Despite the belief that DSIT tools will improve quality and safety, their cost and potential for unintended effects support the need for rigorous evaluations demonstrating positive effects on care and outcomes. These evaluations will ensure that resources are best allocated and that the promise of new technologies is achieved.

### Table 4. Resource Consumption Within 60 Days of the Index Visit by Site and Randomization Status in 1760 Patients*

<table>
<thead>
<tr>
<th>Category</th>
<th>Site 1</th>
<th>Site 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coupler (n = 443)</td>
<td>Usual Care (n = 479)</td>
<td>Coupler (n = 418)</td>
</tr>
<tr>
<td>Ambulatory visits</td>
<td>307 (153-659)</td>
<td>299 (153-590)</td>
<td>307 (146-548)</td>
</tr>
<tr>
<td>Laboratory testing</td>
<td>60 (0-264)</td>
<td>31 (0-182)</td>
<td>31 (0-101)</td>
</tr>
<tr>
<td>Diagnostic imaging</td>
<td>29 (0-159)</td>
<td>29 (0-154)</td>
<td>31 (0-128)</td>
</tr>
<tr>
<td>Pharmacy use</td>
<td>187 (54-555)</td>
<td>164 (41-471)</td>
<td>216 (87-462)</td>
</tr>
<tr>
<td>Total</td>
<td>1023 (423-1950)</td>
<td>831 (372-1721)</td>
<td>701 (321-1315)</td>
</tr>
</tbody>
</table>

*Data are given as median (interquartile range) in dollars.

### Table 5. Patient Satisfaction by Site and Randomization Status for 1573 Respondents*

<table>
<thead>
<tr>
<th>Patient Satisfaction Survey Question</th>
<th>Site 1</th>
<th>Site 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coupler (n = 353)</td>
<td>Usual Care (n = 368)</td>
<td>Coupler (n = 428)</td>
</tr>
<tr>
<td>Speed, efficiency, and courtesy during visit (α = .86)†</td>
<td>4.16</td>
<td>4.13</td>
<td>4.17</td>
</tr>
<tr>
<td>Health care provider (α = .90)†</td>
<td>4.45</td>
<td>4.38</td>
<td>4.35</td>
</tr>
<tr>
<td>Personal issues (α = .81)†</td>
<td>4.29</td>
<td>4.21</td>
<td>4.19</td>
</tr>
<tr>
<td>Overall visit assessment (α = .88)†</td>
<td>4.29</td>
<td>4.20</td>
<td>4.26</td>
</tr>
<tr>
<td>Couplers patients only‡</td>
<td>84</td>
<td>NA</td>
<td>19</td>
</tr>
<tr>
<td>Heard about Couplers before visit, %</td>
<td>15.7</td>
<td>NA</td>
<td>10.9</td>
</tr>
<tr>
<td>Rating of Couplers time spent, min</td>
<td>3</td>
<td>NA</td>
<td>3</td>
</tr>
<tr>
<td>Ease of using Couplers program</td>
<td>4</td>
<td>NA</td>
<td>4</td>
</tr>
<tr>
<td>Recommending Couplers to others</td>
<td>4</td>
<td>NA</td>
<td>4</td>
</tr>
</tbody>
</table>

*Data are given as mean score.

†Data are given as median.

Abbreviation: NA, not applicable.

Includes respondents who answered at least 1 item on the survey.

©2005 American Medical Association. All rights reserved.
Accepted for Publication: July 7, 2004.
Correspondence: Harlan M. Krumholz, MD, SM, 333 Cedar St, PO Box 208088, New Haven, CT 06520-8088 (harlan.krumholz@yale.edu).

Author Contributions: The authors had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Financial Disclosure: None.

Funding/Support: This study was funded by contract V549P-5897 as part of IAW Public Law 104-262 and the Veterans Health Care Eligibility Reform Act of 1996 (38 USC 8151-8153).

Role of the Sponsor: TRICARE approved the study design and reviewed the manuscript but was not involved in the conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation of the manuscript. TRICARE approval of the manuscript was not required.

Acknowledgment: We acknowledge the assistance of Colonel Kenneth Hoffman, project manager at TRICARE Management Activity, and the leadership, providers, and staff at Ireland Army Community Hospital and Clinics and Mayport Branch Health Clinic, who worked patiently with our staff to complete this study.

REFERENCES


