

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

Michael Apkon, MD, PhD, MBA; Jennifer A. Mattera, MPH; Zhenqiu 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

Author Affiliations: Yale–New Haven Health (Drs Apkon, Lin, and Krumholz and Ms Mattera) and Section of Cardiovascular Medicine, Department of Medicine (Drs Herrin, Selter, and Krumholz and Ms Rich), Section of Health Policy and Administration, Department of Epidemiology and Public Health (Drs Bradley and Krumholz), Department of Internal Medicine (Dr Gross and Mr Carbone), and Robert Wood Johnson Clinical Scholars Program (Dr Krumholz), Yale University School of Medicine, New Haven, Conn; and American Board of Internal Medicine, Philadelphia, Pa (Dr Holmboe).

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 performance^{1,2} 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 harm^{9,10} 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.

METHODS

A patient-level RCT was conducted to compare the impact of Couplers with that of usual care at 2 military treatment facilities (Ireland Army Community Hospital and Clinics, Fort

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 model¹⁶ specifying 2 examination rooms and 3.5 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 US 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 proportion of opportunities fulfilled in each study group.

Resources consumed within 60 days of the index visit were determined from the Composite Healthcare System in 4 areas: ambulatory visits, laboratory testing, diagnostic imaging, and pharmacy use. We determined dollar values for the first 3 areas from the Centers for Medicare & Medicaid Services' 2003 fee schedule using the relative value unit conversion rate of \$36.7856.¹⁷ Where not coded in the Composite Healthcare System, Current Procedural Terminology codes were assigned for the midlevel service for each visit type. Drug values were based on the generic equivalent (when available), quantities dispensed, and average wholesale prices found in the 2003 Red Book database (Thomson Micromedex, Greenwood Village, Colo).

Patient satisfaction was assessed using a modification of an outpatient satisfaction survey developed by Press Ganey Asso-

ciates Inc (South Bend, Ind)¹⁸ distributed to participants at their index visit in sealed, numbered envelopes to mask participants to the survey during the visit. Follow-up telephone calls were made when surveys were not returned within 10 business days. Patients who did not complete the survey within 30 days were considered "lost to follow-up" for assessing satisfaction.

We assessed provider satisfaction using a structured survey that asked whether they agreed that Couplers had a positive impact on 8 areas: (1) quality of care, (2) medical decision making (including impact on taking histories, conducting physical examinations, formulating diagnoses, and clinical management), (3) other benefits to patients, (4) patient satisfaction (as reported by providers), (5) patient-provider interaction, (6) time required for patient care, (7) quality of the medical knowledge base, and (8) software design and user interface.

SAMPLE SIZE

We calculated the number of patients required to detect differences in fulfillment rates between the Coupler and usual care groups of 5.0% with a power of 80% at $\alpha = .05$. Calculations used a method¹⁹ that accounted for clustering of outcome measures within patients assuming worst-case baseline fulfillment rates of 50%, opportunities found for each patient in a historical sample at the clinics, and an interpatient correlation of fulfillment 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 χ^2 test of homogeneity, stratified by physician and adjusted for clustering by patient.²⁰ 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 χ^2 tests but stratified by clinic given the inadequate number of individual opportunity types per physician to stratify by physician.¹⁹

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 α) for each domain scale and com-

pared 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).

RESULTS

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 ($P < .01$), 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-

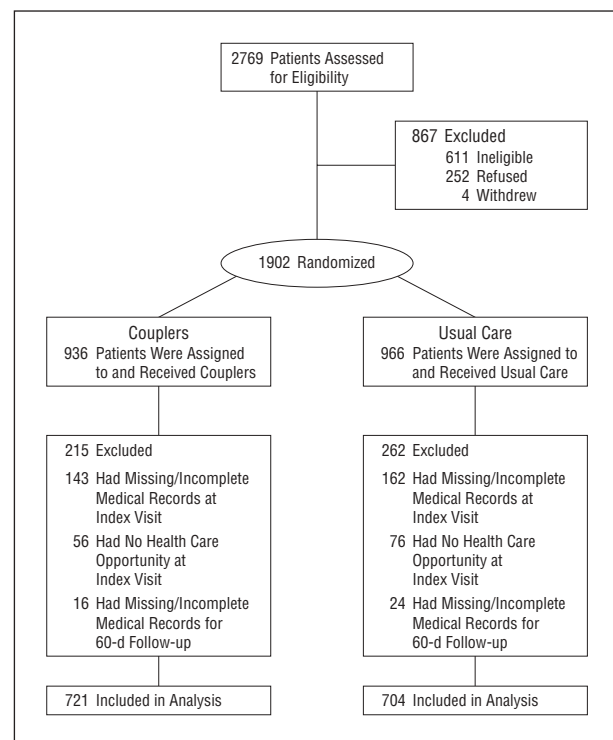


Figure. Patient enrollment and randomization (primary outcome).

Table 1. Characteristics of 1902 Patients by Site and Randomization Status

Characteristic	Site 1		Site 2		Total	
	Coupler (n = 480)	Usual Care (n = 518)	Coupler (n = 456)	Usual Care (n = 448)	Coupler (n = 936)	Usual Care (n = 966)
Age, mean \pm SD, y	34.4 (10.2)	34.4 (9.9)	34.4 (10.6)	36.3 (12.0)	34.4 (10.4)	35.3 (11.0)
Female sex, No. (%)	285 (59.4)	293 (56.6)	308 (67.5)	294 (65.6)	593 (63.4)	587 (60.8)
Military status, No. (%)						
Active duty	215 (44.8)	278 (53.7)	146 (32.0)	147 (32.8)	361 (38.6)	425 (44.0)
Beneficiary	244 (50.8)	225 (43.4)	283 (62.1)	265 (59.2)	527 (56.3)	490 (50.7)
Reserve	1 (0.2)	0	0	0	1 (0.1)	0
Retired	20 (4.2)	15 (2.9)	27 (5.9)	36 (8.0)	47 (5.0)	51 (5.3)
Visit type, No. (%)						
Acute	135 (28.1)	179 (34.6)	248 (54.4)	237 (52.9)	383 (40.9)	416 (43.1)
Established	27 (5.6)	14 (2.7)	20 (4.4)	13 (2.9)	47 (5.0)	27 (2.8)
Routine	228 (47.5)	233 (45.0)	137 (30.0)	142 (31.7)	365 (39.0)	375 (38.8)
Wellness	76 (15.8)	84 (16.2)	50 (11.0)	55 (12.3)	126 (13.5)	139 (14.4)
Other	14 (2.9)	8 (1.5)	1 (0.2)	1 (0.2)	15 (1.6)	9 (0.9)
Health care opportunities						
Screening/prevention						
Patients with ≥ 1 , No. (%)	337 (70.2)	342 (66.0)	350 (76.8)	320 (71.4)	687 (73.4)	662 (68.5)
No. of opportunities/patient, mean	2.14	1.95	2.30	2.17	2.22	2.05
Acute/chronic						
Patients with ≥ 1 , No. (%)	116 (24.2)	127 (24.5)	128 (28.1)	112 (25.0)	244 (26.1)	239 (24.7)
No./patient, mean	0.29	0.31	0.35	0.28	0.32	0.29
Total						
Patients with ≥ 1 , No. (%)	352 (73.3)	358 (69.1)	369 (80.9)	346 (77.2)	721 (77.0)	704 (72.9)
No./patient, mean	2.43	2.26	2.65	2.44	2.54	2.35

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 (33.9% vs 30.7%; $P = .12$) (**Table 2**). The hierarchical multivariable analysis (adjusting for age, sex, military status, visit type, opportunity type, and site) also found no significant effect of assignment to the Coupler group on the probability of health care opportunity fulfillment (odds ratio for opportunities being fulfilled in the Coupler group

compared with the usual-care group: 1.14; 95% confidence interval, 0.95-1.38; $P = .16$) (**Table 3**).

Supplemental Analyses

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).

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 clinic usual-care group (42.5%) and

Table 2. Patient Health Care Opportunities Fulfilled Within 60 Days of the Index Visit by Randomization Status and Site*

Opportunity Type	Site 1		Site 2		Total		P Value
	Coupler (n = 1166)	Usual Care (n = 1171)	Coupler (n = 1208)	Usual Care (n = 1094)	Coupler (n = 2374)	Usual Care (n = 2265)	
Screening/prevention							
Alcohol screening	13/37 (35.1)	3/33 (9.1)	38/42 (90.5)	33/35 (94.3)	51/79 (64.6)	36/68 (52.9)	.07†
Breast cancer	2/4 (50.0)	3/3 (100.0)	1/7 (14.3)	1/9 (11.1)	3/11 (27.3)	4/12 (33.3)	.43
Cervical cancer	6/33 (18.2)	4/37 (10.8)	20/62 (32.3)	18/61 (29.5)	26/95 (27.4)	22/98 (22.4)	.47
Chlamydia	11/31 (35.5)	13/29 (44.8)	11/42 (26.2)	6/35 (17.1)	22/73 (30.1)	19/64 (29.7)	.90
Colorectal cancer	3/21 (14.3)	1/24 (4.2)	1/11 (9.1)	1/34 (2.9)	4/32 (12.5)	2/58 (3.4)	.15
Depression	37/266 (13.9)	30/282 (10.6)	127/156 (81.4)	125/137 (91.2)	164/422 (38.9)	155/419 (37.0)	.58
Dietary counseling	44/235 (18.7)	41/221 (18.6)	105/258 (40.7)	67/228 (29.4)	149/493 (30.2)	108/449 (24.1)	.04
Exercise counseling	49/245 (20.0)	42/229 (18.3)	108/264 (40.9)	67/233 (28.8)	157/509 (30.8)	109/462 (23.6)	.01
Lipid	7/20 (35.0)	11/25 (44.0)	6/29 (20.7)	7/23 (30.4)	13/49 (26.5)	18/48 (37.5)	.32
Pneumococcal vaccine	1/16 (6.3)	0/21 (0.0)	0/45 (0.0)	0/51 (0.0)	1/61 (1.6)	0/72 (0.0)	.25
Smoking/advice to quit	30/105 (28.6)	24/101 (23.8)	62/104 (59.6)	77/99 (77.8)	92/209 (44.0)	101/200 (50.5)	.14
Smoking screening	13/13 (100)	5/7 (71.4)	27/28 (96.4)	24/26 (92.3)	40/41 (97.6)	29/33 (87.9)	.08†
Subtotal	216/1026 (21.1)	177/1012 (17.5)	506/1048 (48.3)	426/971 (43.9)	722/2074 (34.8)	603/1983 (30.4)	.03
Acute/chronic							
Asthma	3/5 (60.0)	4/10 (40.0)	9/13 (69.2)	4/6 (66.7)	12/18 (66.7)	8/16 (50.0)	.57
Back pain imaging	2/2 (100)	1/1 (100)	2/2 (100)	1/1 (100)	4/4 (100.0)	2/2 (100)	NA
Back pain treatment	0/2	1/1 (100)	0/2	1/1 (100)	0/4	2/2 (100)	.05
Diabetes—ACE inhibitor	0/0	1/1 (100)	0/2	0/0	0/2	1/1 (100)	NA
Diabetes—eye examination	1/6 (16.7)	2/10 (20.0)	1/9 (11.1)	1/6 (16.7)	2/15 (13.3)	3/16 (18.8)	.75
Diabetes—hypertension	0/0	1/1 (100)	2/2 (100)	0/0	2/2 (100)	1/1 (100)	NA
Diabetes—glycosylated hemoglobin	0/1	0/2	3/5 (60.0)	1/1 (100)	3/6 (50.0)	1/3 (33.3)	.48
GERD	6/55 (10.9)	6/48 (12.5)	16/83 (19.3)	13/66 (19.7)	22/138 (15.9)	19/114 (16.7)	.85
Hypertension	3/3 (100)	1/3 (33.3)	4/4 (100)	2/4 (50.0)	7/7 (100)	3/7 (42.9)	.03†
Lipid abnormalities	8/48 (16.7)	5/53 (9.4)	4/18 (22.2)	6/16 (37.5)	12/66 (18.2)	11/69 (15.9)	.81
Rhinosinusitis	0/0	0/0	2/3 (66.7)	1/1 (100)	2/3 (66.7)	1/1 (100)	.56
Upper respiratory tract infection	10/18 (55.6)	24/29 (82.8)	7/17 (41.2)	16/21 (76.2)	17/35 (48.6)	40/50 (80.0)	.01
Subtotal	33/140 (23.6)	46/159 (28.9)	50/160 (31.3)	46/123 (37.4)	83/300 (27.7)	92/282 (32.6)	.26
Total	249/1166 (21.4)	223/1171 (19.0)	556/1208 (46.0)	472/1094 (43.1)	805/2374 (33.9)	695/2265 (30.7)	.12

Abbreviations: ACE, angiotensin-converting enzyme; GERD, gastroesophageal reflux disease; NA, not available.

*Data are given as number/total number (percentage).

†Test of homogeneity across sites; $P < .05$.

Table 3. Hierarchical Multivariable Analyses of 4639 Opportunities in 1425 Patients

Category	Health Care Opportunity Fulfillment	
	Odds Ratio (95% CI)	P Value
Coupler vs usual care	1.14 (0.95-1.38)	.16
Opportunity type		
Alcohol screening	0.37 (0.19-0.71)	.003
Asthma	0.33 (0.13-0.87)	.03
Back pain imaging	*	NA
Back pain treatment	0.09 (0.01-0.63)	.02
Breast cancer screening	0.04 (0.01-0.13)	<.001
Cervical cancer screening	0.05 (0.03-0.09)	<.001
Chlamydia screening	0.09 (0.05-0.18)	<.001
Colorectal cancer screening	0.01 (0.00-0.02)	<.001
Diabetes—ACE inhibitor	0.05 (0.00-0.71)	.03
Diabetes—eye examination	0.03 (0.01-0.09)	<.001
Diabetes—hypertension	*	NA
Diabetes—glycosylated hemoglobin	0.13 (0.02-0.67)	.02
Depression screening	0.16 (0.09-0.27)	<.001
Dietary counseling	0.07 (0.04-0.12)	<.001
Exercise counseling	0.07 (0.04-0.12)	<.001
GERD	0.03 (0.02-0.06)	<.001
Hypertension	0.50 (0.12-2.13)	.35
Lipid abnormalities	0.04 (0.02-0.09)	<.001
Lipid screening	0.06 (0.03-0.13)	<.001
Pneumococcal vaccination	0.00 (0.00-0.01)	<.001
Rhinosinusitis	0.37 (0.03-4.58)	.44
Smoking advice	0.18 (0.10-0.32)	<.001
Tobacco screening	3.28 (1.09-9.88)	.04
Upper respiratory tract infection	Referent	NA
Demographics		
Age (≥35 y)	1.34 (1.10-1.63)	.005
Female sex	0.84 (0.64-1.09)	.19
Active military duty	0.94 (0.72-1.24)	.67
Visit type		
Acute	Referent	NA
Established	1.11 (0.66-1.86)	.70
Routine	1.35 (1.08-1.68)	.01
Wellness	3.98 (3.00-5.28)	<.001
Other	0.78 (0.26-2.37)	.66
Site (Fort Knox, Ky)	5.58 (3.90-7.98)	<.001

Abbreviations: ACE, angiotensin-converting enzyme; CI, confidence interval; GERD, gastroesophageal reflux disease; NA, not applicable.

*Predicts fulfillment perfectly or nearly perfectly.

external control clinic (45.7%). The proportion of total opportunities fulfilled in the previous period was also lower at site 1, although the increase during the study period was more modest than at site 2 (10.9%-18.6%).

RESOURCE CONSUMPTION

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 differences were found in the costs (in dollars) associated with ambulatory visits and radiographic evaluation (**Table 4**). However, Coupler group patients used more laboratory and pharmacy resources than usual-care group patients. Aggregate costs across the 4 categories were also higher for the Coupler group, with the median cost per

Coupler group patient being \$100 higher than per usual-care group patient. This estimate did not include direct expenses associated with Coupler use. The average Coupler session took approximately 18 minutes of employee time to coordinate. The logarithmic mean difference between the groups was \$71. Multivariable analysis using logarithmic cost as the outcome showed a significant main effect of treatment, with Coupler group patients using a logarithmic mean difference of \$46 more than usual-care group patients.

PATIENT AND PROVIDER SATISFACTION

Satisfaction surveys were returned by mail or solicited by telephone for 1573 patients (82.7%): 72.2% at site 1 and 92.6% at site 2, with similar response rates between the Coupler and usual-care groups (83.4% vs 82.0%). There were 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%).

COMMENT

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-

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

Category	Site 1		Site 2		Total		P Value
	Coupler (n = 443)	Usual Care (n = 479)	Coupler (n = 418)	Usual Care (n = 419)	Coupler (n = 861)	Usual Care (n = 899)	
Ambulatory visits	307 (153-659)	299 (153-590)	307 (146-548)	265 (138-496)	307 (153-613)	292 (146-541)	.17
Laboratory testing	60 (0-264)	31 (0-182)	31 (0-101)	25 (0-90)	43 (0-144)	31 (0-139)	.04
Diagnostic imaging	29 (0-159)	29 (0-154)	31 (0-128)	0 (0-94)	31 (0-148)	29 (0-127)	.26
Pharmacy use	187 (54-555)	164 (41-471)	216 (87-462)	161 (54-436)	203 (68-495)	164 (50-453)	.03
Total	1023 (423-1950)	831 (372-1721)	701 (321-1315)	647 (304-1154)	789 (375-1654)	698 (340-1530)	.05

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

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

Patient Satisfaction Survey Question	Site 1		Site 2		Total		P Value
	Coupler (n = 353)	Usual Care (n = 368)	Coupler (n = 428)	Usual Care (n = 424)	Coupler (n = 781)	Usual Care (n = 792)	
Speed, efficiency, and courtesy during visit ($\alpha = .86$)†	4.16	4.13	4.17	4.25	4.17	4.19	.23
Health care provider ($\alpha = .96$)†	4.45	4.38	4.35	4.36	4.40	4.37	.82
Personal issues ($\alpha = .81$)†	4.29	4.21	4.19	4.32	4.24	4.27	NA
Overall visit assessment ($\alpha = .88$)†	4.29	4.20	4.26	4.38	4.27	4.30	.74
Couplers patients only‡							
Heard about Couplers before visit, %	84	NA	19	NA	52	NA	NA
Time spent with Couplers program, min	15.7	NA	10.9	NA	13.1	NA	NA
Rating of Couplers time spent	3	NA	3	NA	3	NA	NA
Ease of using Couplers program	4	NA	4	NA	4	NA	NA
Recommending Couplers to others	4	NA	4	NA	4	NA	NA

Abbreviation: NA, not applicable.

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

†Data are given as mean score.

‡Data are given as median.

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.⁴ 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.

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

1. Institute of Medicine Committee on Quality of Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press; 2001.
2. Leaning MS. The new information management and technology strategy of the NHS [editorial]. *BMJ*. 1993;307:217.
3. Hunt D, Haynes R, Hanna S, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA*. 1998;280:1339-1346.
4. Heathfield H, Pitty D, Hanka R. Evaluating information technology in health care: barriers and challenges. *BMJ*. 1998;316:1959-1961.
5. Burkle T, Ammenwerth E, Prokosch HU, Dudeck J. Evaluation of clinical information systems: what can be evaluated and what cannot? *J Eval Clin Pract*. 2001;7:373-385.
6. Littlejohns P, Wyatt JC, Garvican L. Evaluating computerised health information systems: hard lessons still to be learnt. *BMJ*. 2003;326:860-863.
7. Miller RA, Gardner RM; American Medical Informatics Association, Computer-based Patient Record Institute, Medical Library Association, Association of Academic Health Science Libraries, American Health Information Management Association, American Nurses Association. Recommendations for responsible monitoring and regulation of clinical software systems. *J Am Med Inform Assoc*. 1997;4:442-457.
8. Garg AX, Adhikari NKJ, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA*. 2005;293:1223-1238.
9. Goldstein MK, Hoffman BB, Coleman RW, et al. Patient safety in guideline-based decision support for hypertension management: ATHENA DSS. *Proc AMIA Symp*. 2001;00:214-218.
10. Tierney WM, Miller ME, Overhage JM, McDonald CJ. Physician inpatient order writing on microcomputer workstations: effects on resource utilization. *JAMA*. 1993;269:379-383.
11. Johnston ME, Langton KB, Haynes RB, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome: a critical appraisal of research. *Ann Intern Med*. 1994;120:135-142.
12. Institute of Medicine Committee on Quality of Health Care in America. *Leadership by Example: Coordinating Government Roles in Improving Health Care Quality*. Washington, DC: National Academy Press; 2003.
13. Federal Procurement Data System. Available at: <http://www.fpbs.gov/fpdc/FPR2002b.pdf> or <http://sladapp.dior.whs.mil/procurement/Web1999.zip>. Accessed September 14, 2005.
14. Weed LL, Weed L. Opening the black box of clinical judgment: an overview: interview by Abi Berger. *BMJ*. 1999;319:1279.
15. PKC Corp. *Problem Knowledge Couplers: Knowledge-in-Tools*. Burlington, VT: PKC Corp; 2003.
16. "Population health" now military mantra. *US Med*. 1999; 35:1.
17. 68 *Federal Register* 9569 (February 28, 2003) (codified at 42 CFR §410).
18. Drain M. Quality improvement in primary care and the importance of patient perceptions. *J Ambul Care Manage*. 2001;24:30-46.
19. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. London, England: Edward Arnold Publications Ltd; 2000.
20. Donner A, Klar N. Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. *Am J Epidemiol*. 1994;140:279-289.
21. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *N Engl J Med*. 2003;348:2635-2645.
22. Berner ES. *Clinical Decision Support Systems: Theory and Practice*. New York: Springer-Verlag NY Inc; 1999.
23. Berner E, Webster G, Shugerman A, et al. Performance of four computer-based diagnostic systems. *N Engl J Med*. 1994;330:1792-1796.
24. Berner E, Maisiak R, Cobbs C, Taunton O. Effects of a decision support system on physicians' diagnostic performance. *J Am Med Inform Assoc*. 1999;6:420-427.
25. Wellwood J, Johannessen S, Spiegelhalter D. How does computer-aided diagnosis improve the management of acute abdominal pain? *Ann R Coll Surg Engl*. 1992;74:40-46.
26. Berner ES, Maisiak RS. Influence of case and physician characteristics on perceptions of decision support systems. *J Am Med Inform Assoc*. 1999;6:428-434.