

AN EVALUATION OF COLLABORATIVE INTERVENTIONS TO IMPROVE CHRONIC ILLNESS CARE

Framework and Study Design

SHAN CRETIN

American Friends Service Committee

STEPHEN M. SHORTELL

University of California, Berkeley

EMMETT B. KEELER

RAND Health

The authors' dual-purpose evaluation assesses the effectiveness of formal collaboratives in stimulating organizational changes to improve chronic illness care (the chronic care model or CCM). Intervention and comparison sites are compared before and after introduction of the CCM. Multiple data sources are used to measure the degree of implementation, patient-level processes and outcomes, and organizational and team factors associated with success. Despite challenges in timely recruitment of sites and patients, data collection on 37 participating organizations, 22 control sites, and more than 4,000 patients with diabetes, congestive heart failure, asthma, or depression is nearing completion. When analyzed, these data will shed new light on the effectiveness of collaborative improvement methods and the CCM.

Keywords: *chronic disease; methodology; quality improvement; outcome assessment; evaluation design*

CHRONIC ILLNESS CARE AND THE CHRONIC CARE MODEL

More than 40 million Americans have a chronic condition that limits their lives (Institute for Health and Aging UCSF 1996), and as the population ages, the prevalence will increase. Despite significant advances in the management of diabetes, asthma, and other chronic conditions, studies repeatedly find that significant numbers of patients do not receive appropriate care (Harris 2000; Jatulis et al. 1998; Young et al. 2001). Critics argue that systems of care designed to respond to acute episodic illness do not serve the needs of

EVALUATION REVIEW, Vol. 28 No. 1, February 2004 28-51

DOI: 10.1177/0193841X03256298

© 2004 Sage Publications

patients with chronic conditions. The chronic care model (CCM) was developed based on clinical experience and medical evidence to foster improvements in the care of patients with chronic illnesses (Wagner, Austin, et al. 2001; Wagner, Austin, and Von Korff 1996).

The CCM (see Figure 1) integrates diverse elements believed to foster more productive interactions between prepared proactive practice teams and well-informed motivated patients. Provider roles, standards of care, and treatment aims are explicit and evidence based. Care management is linked to a patient registry used for creating reminders, collecting data, scheduling care, and providing performance data to caregivers. Patients are supported through self-management education, participatory goal setting, links to community services, and written care plans. Patients better manage their own care through monitoring, appropriate medication use, and lifestyle choices to enjoy longer more active lives (Calkins, Wagner, and Pacala 1999; Wagner 1998).

In this article, we describe an evaluation of a program that used quality improvement collaboratives to induce the organization, provider, and patient to make the complex changes required by the CCM. A multidisciplinary research team designed a two-pronged evaluation assessing the effectiveness of collaboratives in inducing adoption of the CCM and assessing the effects of implementing the CCM to varying degrees on costs, processes, and outcomes of care. The sections that follow present (a) background on the CCM and collaboratives, (b) interventions being evaluated, (c) evaluation goals, (d) study design, (e) data collection methods, and (f) state of the evaluation.

EVIDENCE BASE FOR THE CCM

The CCM synthesizes evidence from clinical trials of specific practice interventions such as case managers, use of guidelines, and computer-

AUTHORS' NOTE: This work was made possible by the following grants from the Robert Wood Johnson Foundation: No. 034984 and No. 035678. The authors are grateful for the assistance of the Improving Chronic Illness Care Evaluation Team in preparing this article: Steven Asch, Chantal Avila, David Baker, Michael Broder, Julie Brown, Kitty Chan, Sandy Chien, Kathy Dracup, Cecilia Esteves, Cassandra Frankos, Summer Haven, Rosa Elena Garcia, Wendy Hirsch, Geoffrey Joyce, Joan Keesey, David Kurth, Michael Lin, Tom Louis, Carol Mangione, Rita Mangione-Smith, Jill Marsteller, Peter Mendel, Will Nicholas, Lisa Meredith, Marjorie Pearson, Mayde Rosen, Julie Straus, Jurgen Unutzer, Peggy Wallace, Pat Williams, and Shin-Yi Wu. In addition, Donald M. Berwick, and Amy Bonomi and Michael Von Korff provided helpful comments on an earlier draft. Correspondence and requests for reprints should be addressed to Emmett Keeler, RAND Health, 1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138; emmett@rand.org.

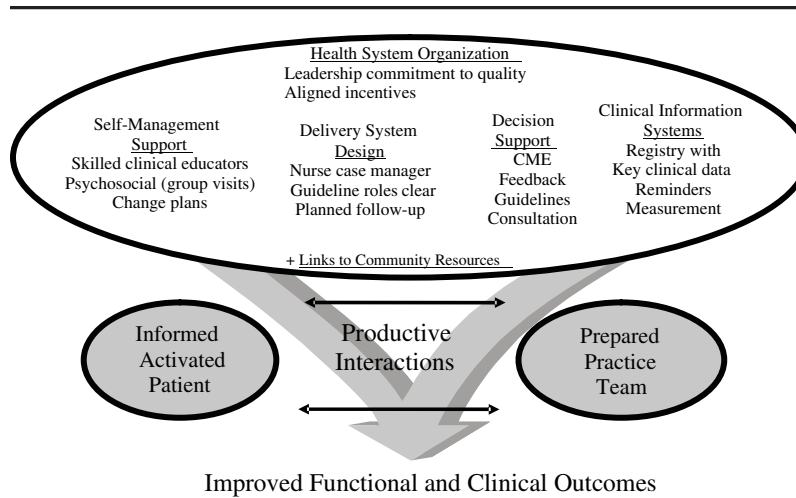


Figure 1: Chronic Care Model
 SOURCE: Adapted from E. H. Wagner, R. E. Glasgow, et al. 2001.
 NOTE: CME = continuing medical education.

assisted reminders (Wagner et al. 1996). Systematic review of hundreds of these studies by the Cochrane Effective Practice and Organization of Care group (<http://www.cochrane.org>) suggests a synergistic effect when individual interventions are combined (Grol and Grimshaw 1999). Trials of integrated strategies of care for diabetes (Renders et al. 2001) and congestive heart failure (Rich 1999) further support the enhanced effectiveness of multifaceted interventions.

Despite its evidence-based origins and intuitive appeal, the CCM has not been evaluated in controlled studies. Individual organizations report better process, outcomes, or costs (Von Korff et al. 1997; Wagner, Austin, and Von Korff 1996) from adopting CCM interventions, but such observational studies are only suggestive. CCM is attractive and plausible, but its effectiveness has not been adequately tested.

In many ways, the CCM and its evaluation are like full-service school evaluations (Shaw and Replogle 1996) and other service coordination interventions (Bickman et al. 2000). Using schools to service more than just educational needs of the students in a less fragmented way is also a plausible idea and shares several grounding ideas with the CCM: a holistic approach to the beneficiary, an attempt to get beneficiaries and their families involved in

improving outcomes, a focus on improving outcomes rather than on procedures, shifts in the roles of staff, and the need to foster community linkages. Some evaluation challenges also are shared: the complexity and variability of the programs call for multiple research methods and make finding out what happened using qualitative methods an important part of evaluation. Schools operate in a more politicized environment than do health organizations, which among other consequences makes it difficult to set up control groups. But in both settings, evaluators must measure and understand the community and organizational context and its changes to understand the degree of success (Cheadle et al. 1998).

COLLABORATIVE METHOD OF QUALITY IMPROVEMENT

Since 1996, the Institute for Healthcare Improvement (IHI) has sponsored Breakthrough Series Collaboratives to foster quality improvement (Gordon et al. 1996). In collaboratives, teams from many organizations work together on a specific problem, guided by evidence-based change ideas, faculty knowledgeable about the problem, and process improvement coaches. Through the collaborative network, teams share effective interventions and strategies for overcoming implementation barriers (Kilo 1998). IHI collaboratives use continuous quality improvement (CQI) methods to make change. CQI is a proactive philosophy of quality management that features (a) multidisciplinary teamwork, (b) empowerment so teams can make immediate process improvements, (c) an iterative scientific approach to problem solving, and (d) ongoing measurement and monitoring. Introduced to health care operations in the late 1980s (Berwick 1989; Laffel and Blumenthal 1989), CQI has been applied to improve processes for individual providers or patients, teams, organizations, and multiorganization systems (Blumenthal and Kilo 1998; Ferlie and Shortell 2001; Laffel and Blumenthal 1989; Wagner, Glasgow, et al. 2001).

PRIOR EVALUATIONS OF COLLABORATIVES

Like the CCM, the collaborative has appealing face validity, but there are few controlled studies of its effectiveness. A review of multisite improvement conducted between 1991 and 1997 found improvements in quality and outcomes of care for the observational before-after designs but little effect for the one randomized study (Shortell, Bennett, and Byck 1998; Solberg et al. 2000). Organizations in IHI-sponsored collaboratives on reducing cesarean section rates (Flamm, Berwick, and Kabcenell 1998) and adverse

drug events (Leape et al. 2000) had varied success in achieving collaborative goals. These mixed effects have been attributed to differences in external environment, organizational culture, available resources, and abilities of organizations to implement interventions (Gordon et al. 1996; Shortell, Bennett, and Byck 1998). The need remains for further investigation of the determinants of successful collaborative improvement (Ferlie and Shortell 2001; Huq and Martin 2000).

Uncertain evidence regarding their effectiveness has not discouraged dissemination of collaboratives around the world, perhaps because of the absence of a more promising alternative (OVretveit et al. 2002). In 1998, the Robert Wood Johnson Foundation National Program on Improving Chronic Illness Care (ICIC) chose to promote adoption of the CCM by cosponsoring three national IHI collaboratives. We evaluated the last two, which focused on diabetes, congestive heart failure, asthma, and depression (Wagner, Glasgow, et al. 2001), together with two collaboratives aimed at improving diabetes care in Washington State led by the ICIC, the Washington State Department of Health, and PRO-West.

THE CHRONIC ILLNESS CARE COLLABORATIVES

Each chronic illness care collaborative imparted three different content areas to the 20 or more participating organizations: an improvement method, the CCM, and condition-specific content (see Figure 2). The improvement approach requires an interdisciplinary team supported by senior organizational leaders, with agreement on aims and a set of related measures for tracking progress. Teams focus initially on a pilot population (about 100 to 300 patients under the care of providers from the collaborative team). They are taught the Plan-Do-Study-Act method for developing and testing process changes on the pilot group and demonstrating improvement with the agreed-upon measures (Berwick 1996, 1998; Kilo 1998; Langley et al. 1996). Senior leaders are asked to support the team, monitor their progress by reviewing monthly reports, and develop a plan for spreading successful process changes throughout the organization.

In these collaboratives, expert CCM and clinical faculty used the CCM and the clinical evidence to develop condition-specific aims (e.g., achieve glycemic control and reduce risk of heart disease in diabetics), measures (e.g., percentage of patients with hemoglobin A1c [HbA1C] > 8.0), and ideas for process changes (e.g., use a registry to track eye and foot exams, blood pressure, lipids, and HbA1C; offer group visits). Organizations also received instruction in the CCM and periodically assessed the degree to which their

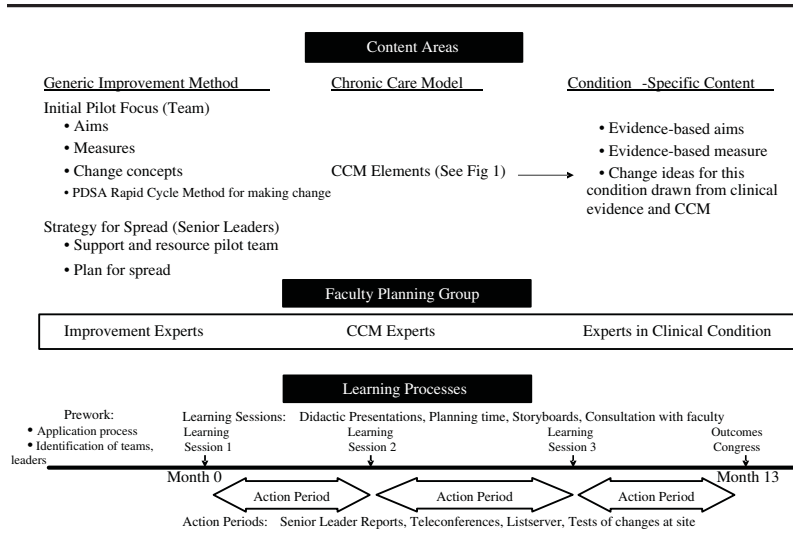


Figure 2: Chronic Illness Care Collaborative Intervention
 NOTE: CCM = chronic care model; PDSA = Plan-Do-Study-Act.

systems reflect ideal practices. Change ideas were organized around the six elements of the CCM, and teams were asked to address all six elements in their efforts to improve care.

The collaboratives used various learning processes. When organizations applied, they were asked to form interdisciplinary teams and review a “prework package,” which introduced the condition-specific aims and measures and also came with a survey for assessing how closely current practices follow the CCM (Assessment of Chronic Illness Care [ACIC] at <http://www.improvingchroniccare.org/tools/acic.html>) (Bonomi et al. 2002). At the three 2-day learning sessions, faculty experts presented lectures and workshops in the improvement method, the CCM, and the specific condition being addressed. Faculty also consulted on specific questions throughout the collaborative. In later learning sessions, participating teams made presentations and storyboards. In the intervening “action periods,” teams developed their interventions by conducting small-scale tests of changes with their pilot patients and providers and shared the results of these tests through the monthly senior leader reports, teleconference calls, and the listserver.

The goal of these collaboratives was to enable participant organizations to make dramatic improvements in patient-care processes and patient outcomes for all of their patients with the target condition. To do so, each team must

develop a locally tailored set of process changes, creating a new system of chronic care based on the principles of the CCM. This system of care must be institutionalized for the pilot population and for the broader target population. The organization must provide resources and a plan for disseminating this new care system.

The changes actually implemented vary from site to site. Within sites, changes involve organization-wide changes (e.g., investment in an information system), changes in processes of care at the clinic level (e.g., patient scheduling procedures or templates for planned visits), and at the individual patient level (e.g., establishing collaborative care goals with each patient).

GOALS OF EVALUATION

The evaluation has two broad, related goals: (a) to gauge the extent to which organizations participating in the collaboratives change their systems for delivering chronic illness care to align with CCM characteristics and (b) to assess the degree to which adopting the CCM improves process and outcomes for patients. We can only study the effect on outcomes of adopting a CCM-based system of care if the organizations enrolled in the collaboratives actually implement the CCM.

Based on self-report, IHI collaboratives do appear capable of catalyzing significant change in most participating organizations (Wagner, Glasgow, et al. 2001). To do so, the Chronic Illness Care Collaborative teams had to make a particularly complex interdependent set of changes that involved many people, departments, and processes within their organizations.

A third major goal is to use naturally occurring variation to investigate environmental, organizational, and team factors associated with success in achieving organization- and patient-level changes. To this end, we hypothesize a multilevel “chain of action” (Whyte 1991) that begins with participating organizations and their environment (see Figure 3). Environmental factors such as pressure from health plans to improve reported performance can push organizations to improve quality with tough competition. Organizations with a supportive culture and a commitment to quality tend to create a positive workplace climate with motivated staff that is able to work together effectively (Ferlie and Shortell 2001; Shortell, Bennett, and Byck 1998). Effective teams can make system changes (such as creating and using a registry) and learn from their experience. These systemic changes in turn affect care processes for individual patients and ultimately for patient outcomes. Finally, success at the patient level feeds back to the team and the organization’s leadership and culture.

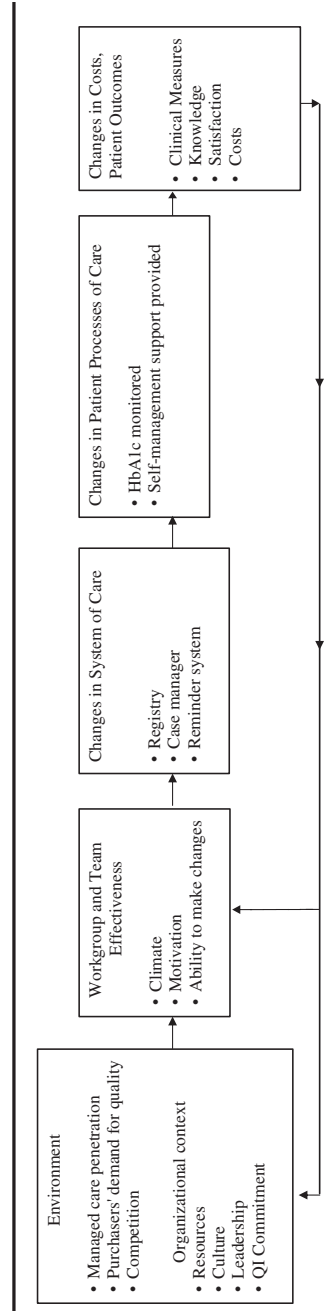


Figure 3: Evaluation Framework: Chain of Effect
 NOTE: QI = quality improvement; HbA1C = hemoglobin A1c.

Measuring all the variables in the chain with standardized assessments allows us to explore the secondary questions of interest:

- Which CCM elements have the greatest impact on patient outcomes?
- Which patients benefit the most from the CCM?
- What is the net cost of implementing the CCM through a collaborative?
- How does participation affect organizational characteristics, team attitudes, and motivation regarding quality?

METHOD

OVERALL STUDY DESIGN

We chose the before-after design with a comparison group as our evaluation design (Cook and Campbell 1979). As shown in Figure 4, we measure patients, care delivery teams, and organizations before and after the introduction of CCM changes, comparing the pilot site with a comparison site that is not immediately affected by the collaborative. In this design, the comparison site helps to control for environmental changes occurring during the collaborative, such as changes in ownership, payment, or competition. Before and after measures allow sites and patients to act as their own controls, reducing the bias from any unmeasured preexisting differences between the pilot and comparison groups.¹

BARRIERS TO RANDOMIZATION IN ORGANIZATIONAL RESEARCH

Our evaluation design did not rely on randomization of organizations, providers, or patients. Although randomized controlled trials are the gold standard in clinical research, in organizational research it is difficult to execute such trials, especially when the intervention requires extensive changes in organizational structure and workflow. Organizations participating in a chronic care collaborative must commit significant resources to the improvement effort. Having made that commitment, many organizations would be unwilling to participate in a randomized study that imposes the additional burden of evaluation while possibly delaying or impeding improvement (through randomization to a delayed intervention or control group).

The collaborative method entails a change strategy that starts with the most enthusiastic providers and then spreads to more resistant staff. Furthermore, the collaborative method recommends that organizations experiment

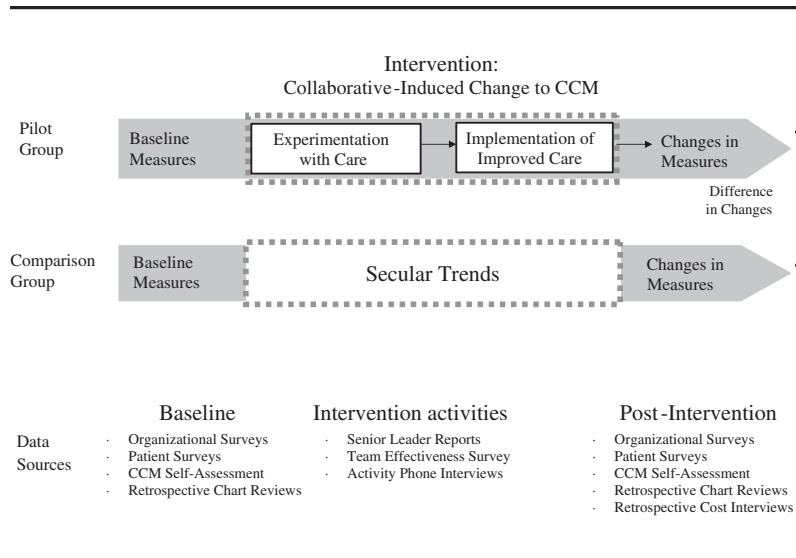


Figure 4: Overview of Study Design

NOTE: CCM = chronic care model.

with and adapt the interventions used with providers and patients. Randomization of providers and patients within an organization to early and late intervention groups with highly structured protocols would substantively change the intervention.

Internal validity is the strength of randomized trials: causality can be clearly established even when the mechanism of action is unknown and potential confounders are not measured. By randomly assigning enough replications to an intervention and a control group, confounders are stochastically balanced, even when they are not identified or measured (Samsa and Matchar 2000; Thorsen and Makela 1999; Weinberger et al. 2001). Even in randomized trials, tight inclusion standards (e.g., 50- to 59-year-old women with newly discovered Stage I breast cancer) or stratification may be used to reduce the chance of actual imbalance between arms of the study. Given the impracticality of randomization and the heterogeneity of organizations, we used before-after pairing and the selection of internal control groups to balance confounders. We will adjust for any remaining imbalances using the many potential confounders measured at the organization and patient level (Cochran 1983).

CHOICE OF COMPARISON GROUPS

Some organizations can provide internal comparison sites such as another clinic in a provider network or another team in a provider setting organized by teams. Despite problems with the intervention changes leaking across staff or spreading prematurely within an organization, we sought internal control groups for three reasons. First, the closer the comparison group matches the intervention group on other factors that affect results, the more valid the conclusion that a true intervention effect exists. Because many changes occur at the organization level (mergers, payment changes), the comparison group must be internal to mirror the effect of those changes. Second, in a voluntary evaluation it is easier to enroll a comparison site from a participating organization with vested interests in the evaluation results than from an organization that is not concerned with the CCM. Finally, the substantial fixed costs per organization of evaluation implies that costs are reduced if the same organization serves as both treatment and control.

TIME TO INITIAL MEASURE AND TO FOLLOW-UP

We planned to time before and after data collection to let teams have full opportunity to implement the interventions. The CCM requires patient registries that teams may need time to develop. Also, although quality improvement teams are told to begin working immediately, the Plan-Do-Study-Act method advises them to start by testing changes on a few patients, learning what works and refining the changes they will ultimately implement. Most teams cannot make process changes affecting a substantial part of their pilot population until 4 to 6 months after the first learning session. So, ideally the first wave of surveys should occur within 4 to 6 months of the initial learning session, with follow-up surveys a year later. For chart reviews, the before period covers 12 months before the first learning session and the after period begins 4 months after the first learning session.

POWER CALCULATION AND SAMPLE SIZES

We calculated the sample sizes needed to detect improvement in patient outcomes across all sites working on the same disease, adjusting the effective sample sizes to account for assumed correlation between variables within a site. We conservatively assumed that adjusting for the baseline value of an

outcome measure would reduce residual variability (based on a longitudinal correlation r of .5), but the effect of other covariates was negligible. Suppose an item has mean 0.3 and variance $\text{Var}(e) = 0.21$, and 100 pilot, 100 control patients finish the study at each site. Then, the variance of the estimated “difference in differences” is 0.0042 ($= 4(1 - r) \text{Var}(e)/100$), producing a standard deviation of 0.065 for a site. So at 80% power and 5% two-sided significance, we can detect $2.8 \times 0.065 = 18\%$ change within a site.

Based on information provided us, the site-specific random effect on the change score is $\text{Var}(s) = 0.0049$, producing a standard deviation of $SD(s) = 0.07$. This produces an intraclass correlation of .0228 ($= 0.0049 / (0.0049 + 0.21)$). So the variance of the estimate pooled over 10 sites is 0.00091 ($= 0.0049/10 + .0042/10$), producing a standard deviation of 0.03. Therefore, the study is powered to detect a pooled $2.8 \times .03 = 8.4\%$ change.

Under the foregoing assumptions, the optimal design for pooled effects uses a number of pilot patients per site equal to $(S \times \text{Var}(s) / \text{Var}(e))$, where $S =$ fixed costs per site/(costs per patient). We assumed that $S = 50$, and $\text{Var}(s) / \text{Var}(e) = 40$, producing an optimal 90 ($= 2 \times 45$) patients per site. With our per-disease budget of \$1 million and a cost per patient of \$400, this leads to $\$20,000 + 90 \times \$400 = \$56,000$ per site, or 18 sites per disease.

DATA SOURCES AND INSTRUMENTS

The evaluation independently assesses the degree to which the CCM was implemented and measures changes in process of care, patient health status, patient satisfaction, provider attitudes, and costs. The chain of action in Figure 3 underscores the need to measure and adjust for organization, team, and patient-level variables in doing so.

Table 1 illustrates selected measures for diabetes (including measures applied to all diseases) and possible data sources. Complete tables of this type for each disease, along with Figure 3, guided development of the surveys and chart review instruments. The amount of information collected from each site and each patient was weighed against the costs, including participant burden and reduced participation rates, if the evaluation tasks are too demanding. All surveys and references to their item sources are available at <http://www.rand.org/health/ICICE>.

Organizational surveys. Brief surveys to capture organization-, workgroup-, and team-level measures were constructed, where possible, using scales and items from previous organizational research in health care and other settings. Surveys were tailored to senior leaders, quality directors,

TABLE 1: Data Collection Strategy for Diabetes

Measure	Data Sources					
	ACIC (CCM Fidelity)	Senior Leader Report and Recap Interview	Cost Interview	Organization Survey	Patient Survey	Chart Review
Organizational and environmental characteristics				X		
Group (size, composition by specialty, compensation)				X		
Clinician and team characteristics (RN, MD, specialty, training, compensation)				X		
Coordination of care between primary and specialty care (involvement of RNs, educators, care managers)	X	X		X	X	
Access to specialists support and resources (direct, preauthorization, mandatory referral for high risk)	X			X	X	X
Organizational culture, leadership, resources, quality orientation				X		
Team characteristics				X		
Workgroup motivation, climate				X		
Team effectiveness				X		
Intervention characteristics						
Baseline and postintervention assessment of CCM fidelity	X					
What changes sites made					X	
Organizational and MIS elements of CCM	X	X		X		
Process of care for diabetes						
Confirmed diagnosis of diabetes					X	X
Patient treatments, monitoring, and referrals					X	X
Attendance at diabetes education classes (diet, smoking cessation, alcohol use, lipid management, diabetes self-care)					X	X

Patient motivation, knowledge, self-efficacy beliefs		X	
Does patient perform self blood glucose monitoring and other useful self-care, adhere to other physician instructions		X	
Patient covariates			
Patient background characteristics, insurance (all diseases)		X	X
Type and duration of diabetes		X	X
Comorbidity (presence and severity) and complications		X	X
Outcomes for all diseases			
Mental and physical functioning (MOS SF-12)		X	
Satisfaction with care for disease (CAHPS)		X	
Utilization of care: hospital, ER, visits, drugs		X	
Disability, lost work days, bed days		X	
Outcomes for diabetes			
Glycemic control, diabetes-related symptoms	X	X	X
Diabetes-related quality of life		X	
Costs			
Cost of participation in collaborative and making changes			X
Cost and revenues associated with new care system			X

NOTE: ACIC = Assessment of Chronic Illness Care; CCM = chronic care model; MOS SF-12 = Medical Outcomes Study Short-Form 12; CAHPS = Consumer Assessment of Health Plans Study.

collaborative team members, and other key players whose cooperation was needed to ensure success. In both the before and after surveys, respondents assessed their organizations' commitment to quality improvement by using a modified Baldrige Quality Improvement Assessment (Shortell et al. 1995), and they also assessed their perceptions of the chronic care collaborative, motivation, and workgroup climate. In the first survey, the organizational culture (a stable trait), experience with quality improvement, and the importance put on improving chronic illness care were assessed. In addition, a key contact completed an organizational characteristics instrument that addressed managed care penetration, demand by purchasers for evidence of quality improvement, competition among local providers, financial position, and stability of executive and clinical leadership. In the after survey, team members also were asked to rate team effectiveness (Baker 1996).

ACIC surveys. Jointly filled out by each participating team at the beginning and end of a collaborative, the ACIC asks for subjective ratings for two to five items in each of the six elements of the CCM. The instrument generates summary scores that assess how closely an organization's structures and processes reflect the CCM ideal.

Senior leader reports and storyboards. Using a uniform format, collaborative sites reported their progress monthly to organization leaders. These senior leader reports described the change activities conducted that month and graphed the site's performance with pilot patients on a collaboratively chosen set of indicators. At each learning session, teams also presented storyboards describing their progress in more detail. These reports and storyboards were forwarded to IHI and the evaluation team. We developed a coding tree and decision rules for categorizing reported changes into specific CCM elements, and we used it to code change activities from all reports or storyboards. NVivo is used to generate a draft log of all CCM change activities for each site.

Team leader recap interviews. We mailed a log designed to clarify the sites' intervention activities to each site before a telephone interview. We also asked about the major successes, major barriers in implementing the CCM efforts and how these were addressed, continuation and spread of the CCM efforts, and procedures for writing and reviewing the senior leader reports. The site contacts have been enthusiastic and forthcoming in discussing their intervention experiences.

Cost interviews. Participation in a collaborative incurs immediate costs associated with learning how to make improvements and incremental costs of testing, refining, and making changes. Later there may be ongoing changes in costs and revenues associated with providing care. Based on the coded activities and recap interview, site-specific data collection worksheets were developed to estimate these costs. After sending the worksheet to the site contact, we conducted a 30-minute call reviewing reimbursement, competitive environment, and the reasons for and financial impact of each intervention.

Patient telephone surveys. The 30-minute telephone surveys had to cover many domains. So, we limited the number of questions for each domain, sampling from items and scales used in previous studies of the targeted diseases. These disease-specific items addressed coordination and process of care, knowledge, self-efficacy, patient behaviors, and outcomes. Some questions were modified slightly for clarity, ease of telephone use, and consistency.

The patient surveys have common elements across diseases: general health (assessed by the Medical Outcomes Study Short-Form 12) (Gandek et al. 1998; Jenkinson et al. 1997); satisfaction with provider (from Consumer Assessment of Health Plans Survey) (Crofton, Lubalin, and Darby 1999; Harris-Kojetin et al. 1999; Hays et al. 1999); and questions relating to age, education, race, language, income, and use of health services in the past 6 months (see Table 1). We worked with the MacColl Institute to develop a short comprehensive set of questions to assess patient-related elements of the CCM.

Chart reviews. We used data from patient medical records to assess changes in key processes of care for diabetes, congestive heart failure, and asthma but not for depression.² Teams of physicians and nurses with condition-specific expertise reviewed existing guidelines and quality indicators for each disease, keeping indicators thought likely to vary significantly and be feasible to implement (e.g., for diabetes, HbA1c testing should be performed annually). New CCM-related measures were developed (documentation of counseling, goal setting, action plans, and specific interventions such as group education visits). We developed a computerized tool into which the data elements required to determine eligibility and to score each measure could be entered by the trained abstractors.

RESULTS

ENROLLMENT OF SITES AND PATIENTS

Enrolling sites and patients into the study was challenging for several reasons. Fewer sites participated in the first IHI collaborative than expected. The second IHI collaborative was composed mainly of Bureau of Primary Health Care community health centers. These centers, often strapped for resources and staff, were reluctant to divert energy from improving care to participation in the evaluation. Some feared their patients would be upset by telephone calls from an unfamiliar source. Few had enough patients to provide both a pilot and comparison sample of adequate size or had the ability to separate pilot and comparison patients. Efforts to get other community health centers as external comparison groups failed, despite help from the Bureau of Primary Health Care.

Our actual site enrollment is shown in Table 2. Even after adding selected sites from the two Washington State Diabetes Collaboratives, we ended up with 37 sites completing the study, of which 22 had comparison groups, instead of the planned 40 fully participating sites. Several participating sites discovered that their pilot population was smaller than expected. Patient enrollment was dependent on the consent process approved by each sites institutional review board (IRB). When we were able to contact patients directly, patient cooperation rates were 85%. When sites had to contact patients first to obtain oral or written consent, the cooperation rates fell to 61% and 39%, respectively (Nelson et al. 2002).

We have about 4,000 patients, which is about half of those initially planned. Diabetes has a larger sample than planned, and the sample in congestive heart failure and asthma is still large enough to detect improvements on a scaled variable of 14% at the site level and 10% pooled. We did not enroll enough sites and patients for the planned analysis for depression and instead will rely primarily on senior leader reports and team leader recap interviews.

TIMING OF SURVEYS

Patients' surveys (which cover current attitudes and care for the preceding 6 months) cannot begin until the sites agree to be in the evaluation, complete the IRB review, and provide a sampling frame of patients with contact information. In the IHI collaboratives, the majority of sites signed up shortly before the first learning session, and it took time to get site agreement. Many

TABLE 2: Sites and Patients Participating in the Evaluation

Collaborative	Disease	Date of First Learning Session	Number of Sites Finishing Collaborative	Number of Sites in Full Evaluation (organization + patient data)	Number of Sites With Comparison Group	Months to Median First Wave Survey	Number of Patients With First Wave Interview (pilot)	Number of Patients With First Wave Interview (comparison)
IHI 1	Congestive heart failure	May 1999	13	7	6	10	400	449
WSDC I	Diabetes	May 1999	8	5	5	10	338	552
	Diabetes	October 1999	4 ^c	2	2	10	201	121
IHI 2	Asthma	February 2000	26 ^a	12	2	14 ^b	509	189
	Depression	February 2000	21 ^a	4	0	16 ^b	140	0
WSDC II	Diabetes	February 2001	9 ^c	7	7	5	584	526
Total				37	22		2,172	1,837

NOTE: IHI = Institute for Healthcare Improvement; WSDC = Washington State Diabetes Collaborative.

a. Many of these sites were not eligible for an evaluation, either because they were too small or were doing something unusual that did not fit with our evaluation tools or consent requirements.

b. For this collaborative, we will only do one patient and provider survey

c. In these collaboratives, we only recruited sites with more than 200 patients.

IRBs, unused to health services research or interventions in which protocols for patients cannot be specified in advance, needed to be convinced about the value and integrity of the study. In sites requiring advanced consent, site staff had to contact each patient before forwarding their information to the evaluation team. The delay from the first learning session to sending out the surveys lagged from the ideal 4 to 6 months to 10 months in IHI 1 (diabetes and congestive heart failure) and to 15 months in IHI 2 (asthma and depression). As a result of these delays, we scheduled the follow-up telephone call 10 months after the first session in IHI 1, and we dropped the follow-up call altogether in IHI 2.

We identified most Washington State Diabetes Collaborative II (WSDC II) sites 2 months prior to the first learning session, allowing us to complete IRB and consent processes and begin surveys shortly after the collaborative began. In all collaboratives studied, we will conduct the chart review as planned, giving us true before and after data.

ANALYSIS PLANS

We will focus on comparing changes in clinical process and patient outcomes between pilot and control sites. We are studying both process and outcome, because even though better outcomes are the ultimate goal of medical care, they can only be achieved through better process and are not as timely or sensitive a measure of quality improvement as is process (Brook, McGlynn, and Cleary 1996). Viewing the collaborative as a way of inducing changes down the chain of effect in Figure 3, we are evaluating all the links in this logic model (from changes in the organization to better patient education and clinical support to knowledge of self-management to self-management behavior to better outcomes). Validating a sensible theory of action reinforces any of our black box conclusions (Bickman 2000). Our evaluation is powered to assess changes in outcomes at the most aggregate level, but we expect to see differences in process in subgroups that will let us explore site characteristics that lead to more change.

In our primary comparison of pilot and control sites, we simply use an indicator for pilot to estimate how well the collaboratives on average improved outcomes. However, data on the number and depth of organizational changes implemented over the course of the collaborative by each pilot site are available and have been scaled. Secondary analyses will be based on that scale to see if greater organizational change leads to greater changes in patient outcomes. In addition, the control sites vary in their degree of organizational separation from the intervention sites (physical distance, shared

staff, and leadership), and we investigate if control sites that are closer to intervention sites benefit more from bleed of the intervention.

The collaboratives have promoted clinical processes that, based on solid evidence, lead to better outcomes (Wagner, Austin, and Von Korff 1996). Because control-site patients are not randomized and are not a perfect match for the pilot sites, a valid analysis requires adjusting for preexisting characteristics such as insurance status, income, education, and baseline duration and severity of the disease (Cochran 1983). To allow for the clustering of process and outcomes within sites (patients are treated by physicians who are members of sites), we use hierarchical regression models (Proc Mixed for measured outcomes, Proc GenMod for binary outcomes) (SAS Institute 1999).

CONCLUSIONS

Our analysis must deal with three further limitations. First, we only have organizations and people who want to improve and have agreed to be evaluated. We have some public information from the collaborative to see if sites that declined to be evaluated differ from those evaluated, and we call them to discuss their successes and barriers briefly. Still, results may not generalize to organizations that lack the will or resources to invest in improvement and in evaluation. Second, the pilot groups volunteer or are chosen to launch the improvement of chronic care, so they may differ from the comparison groups. At our suggestion, many organizations picked the patients at sites to which they wanted to spread the intervention as the comparison group, which should limit differences in motivation. Third, the intervention is at the team level, not the patient level (Whiting-O'Keefe, Henke, and Simborg 1984). Although we have reasonable power for the question of if the collaborative and the CCM worked, tests of which type of organization, context, or intervention works best will be constrained by the limited number of sites.

Despite the limitations, we have the richest set of panel data ever collected on collaborative efforts to implement the CCM, along with process and outcome data on thousands of patients with chronic illness. We have collected extensive data on the potential confounders of differences in outcomes between the treatment and comparison groups that can be used to adjust the results. Overall, we should be able to learn if the collaborative method worked to stimulate change and if the induced change led to better process and outcomes, and we should be able to identify some factors associated with an organization's ability to implement change and improve care.

NOTES

1. Because of the stability of many organizational factors and patients health and knowledge, before and after comparisons also can be efficient. Let $Y_{jt} = \alpha_j + \epsilon_{jt}$ be the value of the measure for unit j at time t , where the fixed effects α_j have variance $V(\alpha)$ and independent error ϵ_{jt} has variance $V(\epsilon)$. Then the correlation of Y over time, $r = V(\alpha)/(V(\alpha) + V(\epsilon))$, the variance of final measures $Y_1 = V(\alpha) + V(\epsilon)$ and the changes $Y_1 - Y_0$ have variance $2V(\epsilon) = 2(1 - r)(V(\alpha) + V(\epsilon))$. This is smaller than the variance of final outcomes if $r > .5$.

2. The primary care medical record was not considered to be as rich a source of information for depression as it was for the other conditions. Heightened concerns about privacy, IRB reviews, and patient consent when mental health diagnoses are involved also contributed to the decision not to seek access to medical records for patients with depression.

REFERENCES

- Baker, R. 1996. *Your experience in the breakthrough series: An assessment of the experience of collaborative participants in carrying out change in the IHI breakthrough series. Waits and Delays Post-Collaborative Survey*. N.p.
- Berwick, D. M. 1989. Continuous improvement as an ideal in health care. *New England Journal of Medicine* 320 (1): 53-56.
- Berwick, D. M. 1996. Quality comes home. *Annals of Internal Medicine* 125 (10): 839-43.
- Berwick, D. M. 1998. Developing and testing changes in delivery of care. *Annals of Internal Medicine* 128 (8): 651-56.
- Bickman, L. 2000. Summing up program theory. In *New directions for evaluation, no. 87*, edited by P. Rogers, R. Hacsí, A. Petro, and T. Huebner, 103-12. San Francisco: Jossey-Bass.
- Bickman, L., E. W. Lambert, A. R. Andrade, and R. V. Penalzoa. 2000. The Fort Bragg continuum of care for children and adolescents: Mental health outcomes over five years. *Journal of Consulting and Clinical Psychology* 68 (4): 710-16.
- Blumenthal, D., and C. M. Kilo. 1998. A report card on continuous quality improvement. *Milbank Quarterly* 76 (4): 625-48.
- Bonomi, A. E., E. H. Wagner, R. E. Glasgow, and M. Von Korff. 2002. Assessment of Chronic Illness Care (ACIC): A practical tool to measure quality improvement. *Health Services Research* 37 (3): 791-820.
- Brook, R. H., E. A. McGlynn, and P. D. Cleary. 1996. Measuring quality of care: Part 2 of 6. *New England Journal of Medicine* 335 (13): 966-70.
- Calkins, C. B., E. H. Wagner, and J. Pacala. 1999. *New ways to care for older people: Building systems on evidence*. New York: Springer.
- Cheadle, A., E. Wagner, C. Anderman, M. Walls, C. McBride, M. A. Bell, R. F. Catalano, and E. Pettigrew. 1998. Measuring community mobilization in the Seattle Minority Youth Health Project. *Evaluation Review* 22 (6): 699-716.
- Cochran, W. G. 1983. *Planning and analysis of observational studies*. New York: John Wiley.
- Cook, T. H., and D. T. Campbell. 1979. *Quasi-experimentation*. New York: Houghton Mifflin.
- Crofton, C., J. S. Lubalin, and C. Darby. 1999. Consumer Assessment of Health Plans Study (CAHPS). Foreword. *Medical Care* 37 (3 Suppl): MS1-9.

- Ferlie, E. B., and S. M. Shortell. 2001. Improving the quality of health care in the United Kingdom and the United States: A framework for change. *Milbank Quarterly* 79 (2): 281-315.
- Flamm, B. L., D. M. Berwick, and A. I. Kabcenell. 1998. Reducing cesarean section rates safely: Lessons from a breakthrough series collaborative. *Birth* 25 (2): 117-24.
- Gandek, B., J. E. Ware, N. K. Aaronson, G. Apolone, J. B. Bjorner, J. E. Brazier, M. Bullinger, S. Kaasa, A. Leplege, L. Prieto, and M. Sullivan. 1998. Cross-validation of item selection and scoring for the SF-12 health survey in nine countries: Results from the IQOLA project. International quality of life assessment. *Journal of Clinical Epidemiology* 51 (11): 1171-78.
- Gordon, P. R., L. Carlson, A. Chessman, M. L. Kundrat, P. S. Morahan, and L. A. Headrick. 1996. A multisite collaborative for the development of interdisciplinary education in continuous improvement for health professions students. *Academic Medicine* 71 (9): 973-78.
- Grol, R., and J. M. Grimshaw. 1999. Evidence-based implementation of evidence-based medicine. *Joint Commission Journal on Quality Improvement* 25 (10): 503-13.
- Harris, M. I. 2000. Health care and health status and outcomes for patients with Type 2 diabetes. *Diabetes Care* 23 (6): 754-58.
- Harris-Kojetin, L. D., F. J. Fowler Jr., J. A. Brown, J. A. Schnaier, and S. F. Sweeny. 1999. The use of cognitive testing to develop and evaluate CAHPS 1.0 core survey items. Consumer assessment of health plans study. *Medical Care* 37 (3 Suppl): MS10-21.
- Hays, R. D., J. A. Shaul, V. S. Williams, J. S. Lubalin, L. D. Harris-Kojetin, S. F. Sweeny, and P. D. Cleary. 1999. Psychometric properties of the CAHPS 1.0 survey measures. Consumer assessment of health plans study. *Medical Care* 37 (3 Suppl): MS22-31.
- Huq, Z., and T. N. Martin. 2000. Workforce cultural factors in TQM/CQI implementation in hospitals. *Health Care Management Review* 25 (3): 80-93.
- Institute for Health and Aging UCSF. 1996. *Chronic care in America: A 21st century challenge*. Princeton, NJ: Robert Wood Johnson Foundation.
- Jatulis, D. E., Y. Y. Meng, R. M. Elashoff, A. L. Schocket, R. M. Evans, A. G. Hasan, and A. P. Legorreta. 1998. Preventive pharmacologic therapy among asthmatics: Five years after publication of guidelines. *Annals of Allergy, Asthma and Immunology* 81 (1): 82-88.
- Jenkinson, C., R. Layte, D. Jenkinson, K. Lawrence, S. Petersen, C. Paice, and J. Stradling. 1997. A shorter form health survey: Can the SF-12 replicate results from the SF-36 in longitudinal studies? *Journal of Public Health and Medicine* 19 (2): 179-86.
- Kilo, C. M. 1998. A framework for collaborative improvement: Lessons from the Institute for Healthcare Improvements breakthrough series. *Quality Management in Health Care* 6 (4): 1-13.
- Laffel, G., and D. Blumenthal. 1989. The case for using industrial quality management science in health care organizations. *Journal of the American Medical Association* 262 (20): 2869-73.
- Langley, G. J., K. M. Nolan, T. W. Nolan, C. L. Norman, and L. P. Provost. 1996. *The improvement guide: A practical approach to enhancing organizational performance*. San Francisco: Jossey-Bass.
- Leape, L. L., A. I. Kabcenell, T. K. Gandhi, P. Carver, T. W. Nolan, and D. M. Berwick. 2000. Reducing adverse drug events: Lessons from a breakthrough series collaborative. *Joint Commission Journal on Quality Improvement* 26 (6): 321-31.
- Nelson, K., R. E. Garcia, J. Brown, C. M. Mangione, T. A. Louis, E. Keeler, and S. Cretin. 2002. Do patient consent procedures affect participation rates in health services research? *Medical Care* 40 (4): 283-88.
- OVretveit, J., P. Bate, P. Cleary, S. Cretin, D. Gustafson, K. McInnes, H. McLeod, T. Molfenter, P. Plsek, G. Robert, S. Shortell, and T. Wilson. 2002. Quality collaboratives: Lessons from research. *Quality and Safety in Health Care* 11 (4): 345-51.

- Renders, C. M., G. D. Valk, S. Griffin, E. H. Wagner, J. T. Eijk, and W. J. J. Assendelft. 2001. Interventions to improve the management of diabetes mellitus in primary care, outpatient and community settings (Cochrane Review). In *Cochrane Library*. Oxford, UK: Update Software.
- Rich, M. W. 1999. Heart failure disease management: A critical review. *Journal of Cardiac Failure* 5 (1): 64-75.
- Samsa, G., and D. Matchar. 2000. Can continuous quality improvement be assessed using randomized trials? *Health Services Research* 35 (3): 687-700.
- SAS Institute. 1999. *SAS/STAT users guide, Version 8*. Cary, NC: Author.
- Shaw, K. M., and E. Repogle. 1996. Challenges in evaluating school-linked services. *Evaluation Review* 20 (4): 424-69.
- Shortell, S. M., C. L. Bennett, and G. R. Byck. 1998. Assessing the impact of continuous quality improvement on clinical practice: What it will take to accelerate progress. *Milbank Quarterly* 76 (4): 593-624.
- Shortell, S. M., J. L. O'Brien, J. M. Carman, R. W. Foster, E. F. X. Hughes, H. Boerstler, and E. J. O'Connor. 1995. Assessing the impact of continuous quality improvement/total quality management: Concept versus implementation. *Health Services Research* 30 (2): 377-401.
- Solberg, L. I., T. E. Kottke, M. L. Brekke, S. Magnan, G. Davidson, C. A. Calomeni, S. A. Conn, G. M. Amundson, and A. F. Nelson. 2000. Failure of a continuous quality improvement intervention to increase the delivery of preventive services. A randomized trial. *Effective Clinical Practice* 3 (3): 105-15.
- Thorsen, T., and M. Makela, eds. 1999. *Changing professional practice: Theory and practice of clinical guidelines implementation*. Copenhagen: Institut for Sundhedsvasen (Danish Institute for Health Services Research and Development).
- Von Korff, M., J. Gruman, J. Schaefer, S. J. Curry, and E. H. Wagner. 1997. Collaborative management of chronic illness. *Annals of Internal Medicine* 127 (12): 1097-102.
- Wagner, E. H. 1998. Chronic disease management: What will it take to improve care for chronic illness? *Effective Clinical Practice* 1 (1): 2-4.
- Wagner, E. H., B. T. Austin, C. Davis, M. Hindmarsh, J. Schaefer, and A. E. Bonomi. 2001. Improving chronic illness care: Translating evidence into action. *Health Affairs (Millwood)* 20 (6): 64-78.
- Wagner, E. H., B. T. Austin, and M. Von Korff. 1996. Organizing care for patients with chronic illness. *Milbank Quarterly* 74 (4): 511-44.
- Wagner, E. H., R. E. Glasgow, C. Davis, A. E. Bonomi, L. Provost, D. McCulloch, P. Carver, and C. Sixta. 2001. Quality improvement in chronic illness care: A collaborative approach. *Joint Commission Journal on Quality Improvement* 27 (2): 63-80.
- Weinberger, M., E. Z. Oddone, W. G. Henderson, D. M. Smith, J. Huey, A. Giobbie-Hurder, and J. R. Feussner. 2001. Multisite randomized controlled trials in health services research: Scientific challenges and operational issues. *Medical Care* 39 (6): 627-34.
- Whiting-O'Keefe, Q. E., C. Henke, and D. W. Simborg. 1984. Choosing the correct unit of analysis in medical care experiments. *Medical Care* 22 (12): 1101-14.
- Whyte, W. F. 1991. *Social theory for action: How individuals and organizations learn to change*. Newbury Park, CA: Sage.
- Young, A. S., R. Klap, C. D. Sherbourne, and K. B. Wells. 2001. The quality of care for depressive and anxiety disorders in the United States. *Archives of General Psychiatry* 58:55-61.

Shan Cretin, Ph.D., M.P.H., is currently regional director of the American Friends Service Committee. Her past research includes studies of clinical practice guidelines and quality improvement in health care.

Stephen M. Shortell, Ph.D., is the dean of the University of California, Berkeley, School of Public Health. His current research deals with strategy, structure, and performance of health care organizations and systems.

Emmett B. Keeler, Ph.D., is a senior mathematician at RAND Corporation. His research includes health economics, cost-effectiveness analysis, and evaluation of health care practices and policies.

Request Permission or Order Reprints Instantly

Interested in copying, sharing, or the repurposing of this article? U.S. copyright law, in most cases, directs you to first get permission from the article's rightsholder before using their content.

To lawfully obtain permission to reuse, or to order reprints of this article quickly and efficiently, click on the "Request Permission/ Order Reprints" link below and follow the instructions. For information on Fair Use limitations of U.S. copyright law, please visit [Stamford University Libraries](#), or for guidelines on Fair Use in the Classroom, please refer to [The Association of American Publishers' \(AAP\)](#).

All information and materials related to SAGE Publications are protected by the copyright laws of the United States and other countries. SAGE Publications and the SAGE logo are registered trademarks of SAGE Publications. Copyright © 2003, Sage Publications, all rights reserved. Mention of other publishers, titles or services may be registered trademarks of their respective companies. Please refer to our user help pages for more details: <http://www.sagepub.com/cc/faq/SageFAQ.htm>

[Request Permissions / Order Reprints](#)