

Effects of a Quality Improvement Collaborative on the Outcome of Care of Patients with HIV Infection: The EQHIV Study

Bruce E. Landon, MD, MBA; Ira B. Wilson, MD, MSc; Keith McInnes, MS; Mary Beth Landrum, PhD; Lisa Hirschhorn, MD, MPH; Peter V. Marsden, PhD; David Gustafson, PhD; and Paul D. Cleary, PhD

Background: Multi-institution collaborative quality improvement programs are a well-established and broadly applicable quality improvement strategy, but there is little systematic assessment their effectiveness.

Objective: To evaluate the effectiveness of a quality improvement collaborative in improving the quality of care for HIV-infected patients.

Design: Controlled pre- and postintervention study.

Setting: Clinics receiving funding from the Ryan White Comprehensive AIDS Resources Emergency Act.

Participants: 44 intervention clinics and 25 control clinics matched by location (urban or rural), region, size, and clinic type.

Measurements: Changes in quality-of-care measures abstracted from medical records of pre- and postintervention samples of patients at each study clinic. Measures examined included use and effectiveness of antiretroviral therapy, screening and prophylaxis, and access to care.

Intervention: A multi-institutional quality improvement collaborative (the "Breakthrough Series").

Results: 9986 patients were studied. Clinical and sociodemographic characteristics of the intervention and control patients were similar ($P > 0.05$). Differences in changes in the quality of care were not statistically significant. The proportion of patients with a suppressed viral load increased by 11 percentage points (from 40.1% to 51.1%) in the intervention group compared with 5.3 percentage points (from 43.6% to 48.8%) in the control group, but this difference was not statistically significant ($P = 0.18$). In addition, rates of appropriate screening tests and prophylaxis did not differ between intervention and control sites.

Limitations: It was not possible to perform a pure randomized trial of the intervention or to assess other measures of quality, such as adherence and satisfaction.

Conclusions: This prospective, matched study of almost 10 000 patients found that a quality improvement collaborative did not significantly affect the quality of care. Additional research is needed to improve methods of teaching and implementing quality improvement programs to achieve better results.

Ann Intern Med. 2004;140:887-896.

For author affiliations, see end of text.

See related article on pp 897-901.

www.annals.org

In the past decade, tremendous improvements have occurred in measuring and monitoring the quality of medical care in the United States. Despite these advances, striking problems with quality persist (1, 2). The quality of care for patients with HIV infection is of particular concern. Substantial evidence shows that obtaining medical services and treatment for patients with HIV infection may lead to longer survival and better quality of life (3, 4), yet serious quality-of-care problems and striking disparities in quality by race and social class have been documented (4-6).

In the 1980s, continuous quality improvement techniques were introduced into health care (7, 8). These strategies emphasize that most quality problems are a result of "system" failings rather than problems with individual practitioners (9). In 1995, the Institute for Healthcare Improvement introduced the concept of the "Breakthrough Series," which brings together health care organizations dedicated to improving the quality of care in particular clinical areas through the application of continuous quality improvement techniques (10). These techniques (known as Plan/Do/Study/Act or PDSA cycles; **Figure 1**) first identify deficiencies in quality, next repeatedly implement small-scale interventions and measure changes, and then refine and expand interventions to improve processes of care (11,

12). Typically, each Breakthrough Series collaborative is composed of 20 to 40 participating health care organizations and a faculty with expertise in the clinical area and quality improvement methods (13). To date, the Institute for Healthcare Improvement has conducted collaboratives with more than 700 teams working on 23 clinical conditions or treatment processes, including improving asthma care and reducing medication errors. Although some evaluations of quality before and after a collaborative support the validity of this approach, only a few limited controlled trials have been conducted (14, 15).

An important source of funding for HIV care is the Ryan White Comprehensive AIDS Resources Emergency (CARE) Act, which is administered by the HIV/AIDS Bureau of the Health Resources and Services Administration. Title III of the CARE Act supports comprehensive primary health care for HIV-infected individuals and currently supports primary care services for more than 150 000 patients receiving care in more than 200 community health centers, hospital-based clinics, and city or county health services (16). In 1999, the Health Resources and Services Administration required all clinical sites that were newly awarded funding under Title III of the CARE Act to participate in a quality improvement collaborative conducted by the In-

Context

Multi-institutional quality improvement collaboratives are popular, but are they effective?

Contribution

This controlled study evaluated an HIV care “Breakthrough Series” program that emphasized provider teams, sessions on quality improvement theory and techniques, and report backs about implementing quality improvement. A review of the medical records of 9986 HIV-infected patients showed no important differences in quality of care (viral load suppression, pneumocystitis prophylaxis, and screening for tuberculosis and hepatitis) between the 44 intervention clinics and the 25 control clinics.

Cautions

Patient adherence and satisfaction were not measured, and some control clinics may have used quality improvement techniques similar to those recommended by the collaborative program.

–The Editors

stitute for Healthcare Improvement. Other sites already receiving Title III funding were also invited to participate. This study evaluates the impact of the collaborative by examining pre- and postimplementation quality-of-care information on samples of patients from both participating and matched nonparticipating clinics.

METHODS

Study Site Selection and Controls

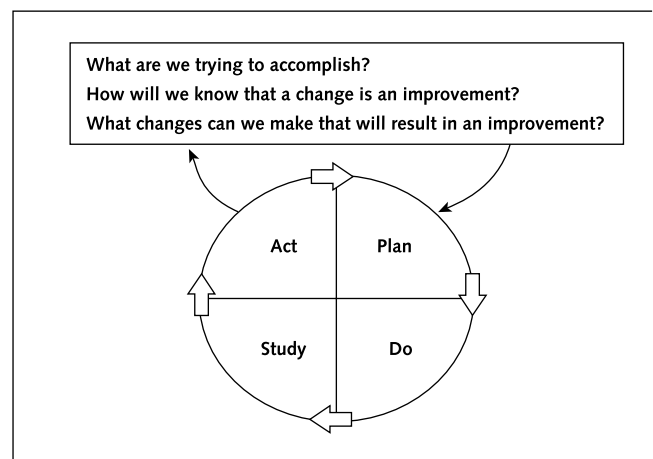
Of the 200 Title III sites in the continental United States in May 2000, we excluded 16 that reported case-loads of fewer than 100 cases per year, 12 that were initially slated to participate in the Breakthrough collaborative but elected not to do so, and 1 that lost its CARE Act funding shortly before the collaborative began. Of the remaining 171 sites, 62 participated in the collaborative. Among these sites, 54 agreed to participate in the study and 44 (including 11 mandatory participants and 33 voluntary participants; 71% of collaborative participants) provided chart review data. Of the 109 nonparticipating sites eligible to be selected as control sites, 65 provided information needed for matching. The potential control sites were matched with intervention sites on the basis of the type of site (community health center, community-based organization, health department, hospital, or university medical center), location (rural or urban), number of locations delivering care, region, and number of patients with active HIV infection. When these criteria were used, 40 sites were selected as potential controls and 37 of them (93%) agreed to participate in the study. Of these, 25 (63% of potential control sites) participated in the chart review portion of the study. The Committee on Human

Studies of Harvard Medical School approved the study protocol.

Quality Improvement Intervention

Each participating clinic selected a team, usually consisting of at least one administrator and one or more clinicians, and a “population of focus” on which the team’s interventions would be tested. Usually, the population of focus consisted of all HIV-infected patients in a particular site, but participants sometimes chose to focus on a subset of patients, such as those under the care of a particular group of clinicians. Originally, the design of the collaborative extended for 12 months and included a “kickoff” meeting and 2 subsequent 2-day meetings called learning sessions. The kickoff learning session included instruction in the theory and practice of quality improvement by identifying problems in HIV care and then introducing the techniques of continuously implementing, measuring, and refining changes (the Plan/Do/Study/Act cycles) (11, 12) to improve the care of HIV-infected patients. Each learning session included additional instruction in quality improvement techniques and breakout sessions that focused on improving specific aspects of care, developing an information infrastructure to track progress, and specific aspects of quality improvement theory. In addition, teams exchanged ideas and presented “storyboards” of their progress to date. At each session, teams reported on activities, methods, and results. Toward the end of the 12-month period, the Health Resources and Services Administration decided to extend the collaborative by 4 months and add a third learning session. Between the sessions (“action periods”), team members implemented concepts and ideas. Each site had access to a collaborative listserv, participated in monthly conference calls with the collaborative faculty, and submitted monthly reports of its improvements, which included charts that tracked the site’s im-

Figure 1. Theoretical construct of continuous quality improvement.



Reproduced from Berwick DM, Nolan TW (11) with permission from the American College of Physicians.

Table 1. Quality of Care Indicators*

Quality Indicator	Eligible Population	Meets Quality Measure
Tuberculosis screening during review period†	All patients	Receipt of a tuberculosis screening test or known to have a positive tuberculosis screening test result in the past
Influenza shot	All patients	Documentation of receipt of an influenza shot during the review period
Hepatitis C status	All patients	Documentation of hepatitis C status during the review period‡
Pap smear	All women	Pap smear performed (or offered and declined) at least once during the review period or note of a Pap smear done elsewhere
Prophylaxis against <i>Pneumocystis carinii</i> pneumonia	All patients with CD4 ⁺ cell counts <0.200 × 10 ⁹ cells/L at any point during the review period	Receipt of prophylaxis against <i>P. carinii</i> pneumonia
Receipt of HAART on last visit for appropriate patients	All patients with CD4 ⁺ cell counts <0.500 × 10 ⁹ cells/L, with viral loads >10 000 copies/mL, or already taking HAART	Receipt of 3-drug HAART regimen
Viral load controlled on last visit (<400 copies/mL or undetectable)	All patients receiving or eligible for HAART	Viral load undetectable or <400 copies/mL
Proportion of patients with increased CD4 cell count	All patients with initial CD4 ⁺ cell counts <0.200 × 10 ⁹ cells/L and at least one other check for CD4 ⁺ cell count	CD4 cell count >0.200 × 10 ⁹ cells/L on last CD4 ⁺ cell count
Visits in 3 or 4 quarters	All patients	Visits in at least 3 or 4 quarters

* HAART = highly active antiretroviral therapy; Pap = Papanicolaou.

† Patients with documentation of a previous tuberculosis screening test were counted as being screened. Those with missing values were counted as not being screened.

‡ Includes those with documentation of a previous positive test result for hepatitis C.

improvements to date in the required key quality measures described in the next section. Detailed descriptions of the Breakthrough Series collaboratives are available elsewhere (10, 17–19).

Quality-of-Care Monitors

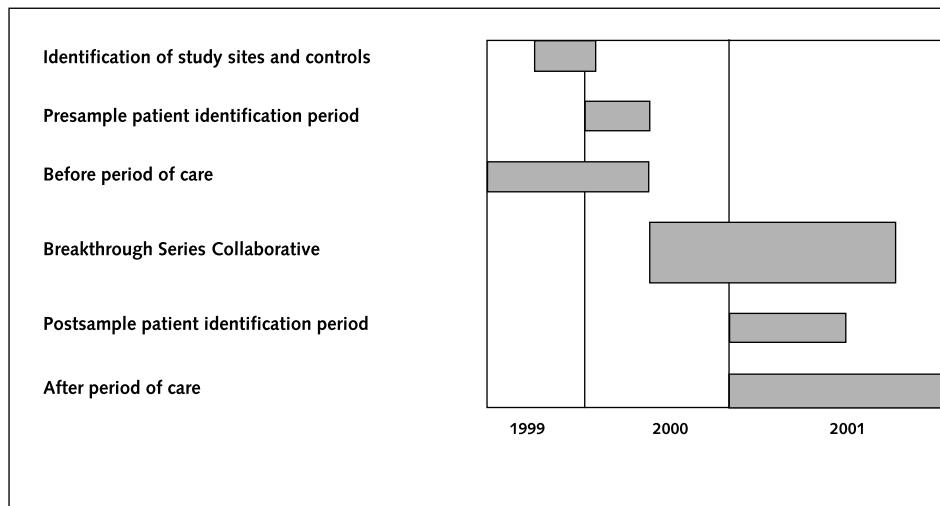
We selected quality-of-care measures (Table 1) to coincide with required and optional quality measures selected by the collaborative faculty as areas for improvement. These measures were selected by the faculty after reviewing the literature to identify areas of quality deficiency in the delivery of HIV care, particularly for underserved populations targeted by the CARE Act. Because of the paramount importance of antiretroviral therapy to the treatment of HIV infection, the faculty focused on measures related to antiretroviral treatment, including the percentage of patients receiving highly active antiretroviral therapy, the percentage of patients with a controlled viral load, and the percentage of patients who received adherence counseling, as required key measures for the collaborative. Measures were then developed on the basis of consensus guidelines appropriate for the period of care (20). Our primary measures were rates of highly active antiretroviral therapy use and control of HIV viral load for appropriate patients. Patients eligible for highly active antiretroviral therapy included those with CD4⁺ cell counts less than 0.350 × 10⁹ cells/L, those with CD4⁺ counts between 0.350 and 0.500 × 10⁹ cells/L and a viral load greater than 5000 copies/mL, all patients with a viral load greater than 30 000 copies/mL, and patients already receiving highly active antiretroviral therapy, as per the guidelines. We also assessed the use of highly active antiretroviral therapy for

those with CD4⁺ counts less than 0.350 × 10⁹ cells/L to reflect recommendations that were published after the end of the collaborative (21). Because of the variability in viral load assays available at the time, viral load was considered controlled if it was undetectable or if the total viral load was less than 400 copies/mL. We also assessed the use of screening and prophylaxis, as well as access to care. The only key measure followed by the collaborative that we could not assess was related to adherence counseling because this information is not reliably available from medical records.

Quality-of-Care Data Collection

To identify pre- and postintervention samples of patients, we requested lists of all HIV-infected patients in care at each of the sites during the 2 time periods (Figure 2). For the first sample, sites were asked to provide encrypted lists of all HIV-infected patients age 18 years or older as of June 2000 seen at the site between 1 January and 30 June 2000. For the second sample, sites were asked to provide a similar list of active patients age 18 years or older as of December 2001 seen at the site between 1 July and 31 December 2001 (the last 6 months of the “post”-study period). We randomly selected 75 patients from each of these lists for the chart review portion of the study. We selected 5 additional patients for each review period to be substituted for cases in which the chart could not be located or was otherwise unavailable. We asked the 11 intervention sites that had focused their improvement efforts on a subset of patients to indicate such patients on their lists, and we randomly selected half of the sample from the population of focus and the remaining half from the other

Figure 2. Design and time line of the Evaluation of Quality Improvement for HIV (EQHIV) Project.



Midway through the process, the collaborative was extended 3 months past its originally scheduled end date.

patients. Each site selected 1 or 2 medical records reviewers (typically nurses). Data were then collected from each sampled patient’s medical record covering a 1-year period of care. The first period was selected to end just as the collaborative began, and the second one extended approximately 6 months beyond the originally scheduled end of the collaborative (2 months past the extended time) to ensure maximal overlap with the end of the collaborative. The data abstracted included sociodemographic information (for example, age and sex), history of HIV-related illnesses, comorbid medical or psychiatric conditions (current substance abuse or psychiatric illness), screening and prophylaxis against HIV-related conditions, number and timing of visits, CD4⁺ cell counts, viral loads, and antiretroviral medications prescribed.

Descriptions of Changes

In addition to the quality-of-care data, we also coded and tracked all of the change initiatives attempted, as noted in the monthly reports submitted by each of the participating clinics. Finally, both control and intervention sites were surveyed at the beginning and the end of the intervention to ascertain their quality improvement activities and environment.

Statistical Analysis

We compared characteristics of the intervention and control sites to each other and to all Title III sites in the continental United States. On the basis of initial power calculations, we designed our study to include 15 intervention clinics and 15 control clinics with a total of 50 patients selected per clinic per round. We estimated power for our 2-level model by using standard formulas for 2 group comparisons in which the total variance of changes in quality was assumed to be equal to the sum of the average within-clinic variance plus the between-clinic variance. Assuming that the standard deviation in changes in

quality across clinics ranged from 7.5% to 15%, the original design would have resulted in 80% power to detect a difference of 13 to 16 percentage points in the rate of change in the main outcome measures. We subsequently decided to increase our sample size by including 44 intervention clinics and 25 control clinics and examined 75 patients per clinic per round to increase our power to detect differences of approximately 10%, which we thought to be clinically significant.

Assessments of mean changes from baseline to follow-up within each of the 2 groups and relative change over time between the 2 groups were examined by using hierarchical regression models that controlled for both patient clinical and sociodemographic characteristics and clinic characteristics. Logistic regression models were used to account for the binary nature of our quality measures. These models included a random intercept and period term to account for correlation among patients within a clinic and allow for differential changes in outcomes across sites. We tested for differences in changes in outcomes between intervention and control sites by using an interaction term between period of measurement (pre- vs. postintervention) and treatment group (intervention vs. control). Patient control variables included age, sex, stage of disease based on lowest recorded CD4⁺ cell count over the period of care, active psychiatric or substance abuse problem, history of HIV-related diagnoses, and other comorbid medical conditions. Because we thought that change might be different for Title III clinics that were required to participate in the program, we compared the effectiveness of the intervention for newly funded clinics and those with ongoing funding. None of the newly funded clinics was new to providing HIV care. All but one had been providing HIV care for 5 or more years. We also stratified clinics at baseline as poorly performing (the bottom 50%) and ex-

amined the effectiveness of the intervention for this group of clinics separately. We also assessed the relative effectiveness of the intervention among patients in the population of focus. Because new guidelines for the initiation of antiretroviral therapy were released in the year after the completion of the collaborative, we also reassessed the performance on the highly active antiretroviral therapy and viral load indicators after restricting the eligible population to those already receiving highly active antiretroviral therapy or those with CD4⁺ cell counts less than 0.350×10^9 cells/L (21). Finally, to test the sensitivity of our conclusions to modeling assumptions, we refit models by using fixed site effects.

Role of the Funding Source

This research was funded by the Agency for Healthcare Research and Quality. The funding source had no role in the design, conduct, or reporting of the study or in the decision to submit the manuscript for publication.

RESULTS

Study Sites

Table 2 reports characteristics of the collaborative participants and control sites. None of the differences between the intervention and control sites was statistically significant, and the sites were representative of Title III clinics nationally. Most clinics were located in urban areas (77% of intervention clinics and 84% of control clinics; $P > 0.2$), and the clinics were well distributed throughout the country. Even after matching, more collaborative clinics were in the South, and more control clinics were in the West, but these differences were not significant ($P = 0.06$). Thirty percent of intervention clinics and 40% of control clinics were community health centers, and approximately 15% were hospital-based clinics ($P > 0.2$).

Clinics were approximately evenly split between large (>400 HIV-infected patients) and small (<400 HIV-infected patients) size, and most were HIV specialty clinics.

Among the 25 control sites, 4 (16%) reported that they had previously participated in a Breakthrough Series collaborative for a different disease and none were currently participating in a different collaborative. As another test of the impact of the collaborative on participating clinics, we interviewed clinic directors from the intervention and control sites after the end of the collaborative. Although both intervention and control clinic directors reported that they had undertaken quality improvement initiatives for HIV care, many more were reported in the intervention clinics. Twenty-seven percent of the intervention clinic directors reported implementing 6 or more initiatives; none of the control clinic directors reported implementing this number of initiatives. Conversely, 64% of control clinic directors reported attempting between 0 and 2 initiatives compared with 27% of intervention clinics.

Participation in the Collaborative by Sites and Types of Interventions Attempted

All clinics attended at least 3 of the 4 learning sessions, and 75% submitted monthly reports 7 or more times (mean, 10.6). All clinics also participated in conference calls and the collaborative listserve, but we have no way of quantifying the level of participation in these activities. We identified 1479 change initiatives attempted at 43 of the intervention clinics that submitted monthly reports, a mean of 33.6 initiatives per clinic. Table 3 reports the number of initiatives tried in specific clinical areas. The most common targets of initiatives were access to care and antiretroviral therapy, with an average of 9.1 and 6.9 identified initiatives per clinic, respectively. Fewer initiatives, however, were either being refined in preparation for im-

Table 2. Description of All Title III Sites and Evaluation of Quality Improvement for HIV Study Intervention and Control Sites in the Continental United States

Variable	All Title III Sites (n = 200)	Intervention Sites (n = 44)	Control Sites (n = 25)	P Value*
Location				
Urban, %	79.3	77.3	84.0	>0.2
Region, %				
South	27.7	36.4	16.0	0.06
West	17.5	6.8	24.0	
Midwest	15.0	15.9	28.0	
Northeast	39.8	40.9	32.0	
Clinic type, %				
Community health center	38.9	29.5	40.0	>0.2
Hospital	11.1	13.6	16.0	
Other†	50.0	56.8	44.0	
HIV-infected patients ± SD, n	623 ± 733	657 ± 695	490 ± 295	0.17
Clinic size				
Small (<400 patients with HIV infection), %	49.0	45.4	44.0	>0.2
HIV specialty clinic				
Yes, %	74.3	81.8	84.0	>0.2

* P value is for testing the differences between intervention and control sites.

† Other includes community-based organizations, university medical centers, and public health clinics.

Table 3. Description of Specific Interventions Attempted by Individual Sites*

Intervention Target	Mean Interventions per Site	Mean Interventions per Site That Were Being Refined or Implemented	Sites with at Least 1 Intervention	Sites with at Least 1 Intervention That Was Being Refined or Implemented
	<i>n</i>		<i>n</i> (%)	
Antiretroviral therapy	6.9	1.1	41 (95)	22 (51)
Screening†	3.5	0.6	39 (91)	18 (42)
Access to care‡	9.1	1.5	42 (98)	27 (63)
Prevention§	1.3	0.1	27 (63)	4 (9)
Women's health	1.7	0.5	27 (63)	16 (37)

* Refers to 43 sites with at least 1 report.

† Screening includes tests such as tuberculosis screening and hepatitis screening.

‡ Access includes visit frequency, continuity of care, annual examinations, clinic hours, ease of appointments, and transportation.

§ Prevention includes vaccinations and prophylaxis against *Pneumocystis carinii* pneumonia.

|| Women's health includes Papanicolaou screening and obstetric and gynecology services.

plementation across the entire clinic or were actually implemented. On average, each of the target areas we examined had one or fewer of these later-stage initiatives per clinic, although more than 50% of clinics reported initiatives at this stage in the areas of antiretroviral care and access to care.

Demographic Characteristics and Clinical History

Overall, 9986 patients (6406 from the intervention clinics and 3580 from the control clinics) were studied, representing 96% of the requested sample. The average number of patients per center in each round was 72 (range, 37 to 77; median, 75). Only a few clinical and sociodemo-

Table 4. Description of the Patient Population by Time Period and Intervention or Control Clinic

Variable	Before Assessment		After Assessment	
	Intervention Clinic (n = 3190)	Control Clinic (n = 1761)	Intervention Clinic (n = 3216)	Control Clinic (n = 1819)
Mean age ± SD, y	39.6 ± 8.7	40.0 ± 8.9	41.5 ± 8.9*	41.7 ± 9.1†
Men, %	66.4	71.3‡	65.8	71.2§
Primary provider, %				
Physician	81.7	80.3	80.0*	79.5
Physician assistant, nurse practitioner, or other	18.4	19.7	21.0	20.5
HIV-related diagnoses ± SD, n	1.31 ± 0.59	1.30 ± 0.58	1.32 ± 0.67	1.21 ± 0.42¶
Any HIV-related diagnosis, %	13.3	13.8	12.9	13.7
<i>Pneumocystis carinii</i> pneumonia, %	3.7	4.9**	3.5	4.2
<i>Candida</i> esophagitis, %	4.3	3.8	5.0	3.5¶
The AIDS wasting syndrome, %	4.6	4.6	4.2	4.6
Comorbid conditions ± SD, n††	1.30 ± 0.59	1.42 ± 0.69‡	1.33 ± 0.57	1.39 ± 0.63
Cardiovascular, %	15.5	12.7‡	16.9	16.2†
Endocrine, %	5.3	5.5	6.0	5.6
Gastrointestinal, %	4.0	5.3	4.1	5.7¶
Lipid, %	12.6	14.2	16.4*	16.6
Psychiatric disorders, %‡‡	30.4	32.0	30.7	35.2§§
Substance abuse, %	15.7	18.0**	14.1	18.2§
Lowest CD4 cell counts, %				
0.0–0.049 × 10 ⁹ cells/L	11.0	8.6	11.0	8.7
0.050–0.199 × 10 ⁹ cells/L	21.3	21.4	20.9	19.0
0.200–0.499 × 10 ⁹ cells/L	43.9	42.8	42.0	45.4
0.500 × 10 ⁹ cells/L	23.9	27.3	26.1	26.9

* P < 0.01 for comparison between column 2 and column 4.

† P < 0.01 for comparison between column 3 and column 5.

‡ P < 0.01 for comparison between column 2 and column 3.

§ P < 0.01 for comparison between column 4 and column 5.

|| HIV-related diagnoses include *Pneumocystis carinii* pneumonia, *Mycobacterium avium* complex infection, *Candida* esophagitis, Kaposi sarcoma, the AIDS wasting syndrome, cytomegalovirus retinitis and enteritis, and HIV encephalopathy.

¶ P < 0.05 for comparison between column 4 and column 5.

** P < 0.05 for comparison between column 2 and column 3.

†† Cardiovascular comorbid conditions include hypertension and coronary artery disease; endocrine comorbid conditions include diabetes; gastrointestinal comorbid conditions include active hepatitis B or active other chronic liver disease; and lipid disorders include elevated cholesterol level or lipodystrophy.

‡‡ Psychiatric disorders include depression, schizophrenia, or bipolar, or anxiety disorder.

§§ P < 0.05 for comparison between column 3 and column 5.

||| Substance abuse includes documented current use of heroin, cocaine, injection drugs, and other illegal drugs or current use of alcohol.

Table 5. Adjusted Quality-of-Care Indicators for Intervention and Control Clinics*

Variable	Before Assessment		After Assessment		Difference		P Value
	Intervention Clinic (n = 3190)	Control Clinic (n = 1761)	Intervention Clinic (n = 3216)	Control Clinic (n = 1819)	Intervention Clinic (n = 6406)	Control Clinic (n = 3580)	
	← % →				percentage points		
Antiretroviral therapy							
Receipt of HAART on last visit for appropriate patients	83.5	80.4	80.5	77.5	-3.0	2.9	>0.2
Viral load controlled	40.7	44.1	51.7	49.5	11.0	5.4	0.18
Screening and prophylaxis							
Tuberculosis screening	52.2	54.2	52.3	52.3	0.1	-2.1	>0.2
Influenza shot	45.8	52.7	53.1	59.5	7.3	6.8	>0.2
Hepatitis C status	84.4	83.5	90.0	89.7	5.5	6.2	>0.2
Pap smear	60.4	66.4	65.0	62.2	4.6	-4.2	0.06
Prophylaxis against <i>Pneumocystis carinii</i> pneumonia	74.7	67.9	75.1	71.4	0.4	3.5	>0.2
Access to care							
Visits in 3 or 4 quarters	68.0	64.7	73.4	67.4	5.4	2.7	>0.2

* Adjusted rates are calculated by using hierarchic logistic regression models after adjustment for patient age, sex, HIV-related diagnoses, comorbid conditions, and lowest CD4 cell count over the review period, as well as for clinic location, region, and type, site size, and specialty status. Adjusted rates represent the predicted value for a patient with average values for each of the control variables. HAART = highly active antiretroviral therapy; Pap = Papanicolaou.

graphic characteristics significantly differed between the intervention and control patients (Table 4). In addition, the pre- and postintervention samples had few significant differences. The average patient was approximately 40 years of age, and about two thirds of the patients were male. Patients in the postintervention period were slightly older ($P < 0.01$ for both the intervention and control group). Eighty percent were cared for primarily by a physician. About 13% of patients had any HIV-related diagnoses; the most common were *Pneumocystis carinii* pneumonia, *Candida* esophagitis, and the AIDS wasting syndrome (each present in approximately 4% to 5% of the sample). Those with any HIV diagnosis had an average of 1.3 HIV-related conditions. Similarly, the patients in the study had 1.3 other comorbid conditions; cardiovascular conditions were most common (present in about 15% of the sample). Approximately 30% of the patients had a psychiatric comorbid condition and 15% were current substance abusers.

Change in Performance: Impact of the Breakthrough Series

The proportion of patients with a suppressed viral load increased by 11 percentage points (from 40.7% to 51.7%) in the intervention group compared with 5.4 percentage points (from 44.1% to 49.5%) in the control group, but this difference was not statistically significant ($P = 0.18$) (Table 5). Rates of highly active antiretroviral therapy use for appropriate patients were high in both the pre- and postintervention periods (approximately 80%), with a slight decrease in the second period. The intervention and control sites did not differ significantly.

Differences for screening and prophylaxis between the pre- and postassessment periods were small for both the intervention and control clinics. Tuberculosis screening tests were conducted for half of the patients, and the difference between the intervention and control sites was

small. Influenza shots and documentation of hepatitis C exposure status increased for both study groups (7.3 percentage points and 6.8 percentage points for influenza shots for the intervention and control clinics, respectively [$P > 0.2$], and 5.5 percentage points and 6.2 percentage points for hepatitis C screening for the intervention and control clinics, respectively [$P > 0.2$]). Papanicolaou smear rates increased by 4.6 percentage points in the intervention clinics from a baseline level of just 60.4% and decreased by 4.2 percentage points in the control clinics from a baseline level of 66.4% ($P = 0.06$).

Finally, both the intervention and control clinics showed a slight increase in the number of patients with visits in at least 3 of the last 4 quarters. From a baseline rate of approximately 65%, the collaborative clinics improved by 5.4 percentage points compared with 2.7 percentage points for the control clinics ($P > 0.2$).

The effectiveness of the intervention did not differ significantly for any of the outcomes for new Title III ($n = 11$) versus ongoing Title III ($n = 33$) clinics, nor was the intervention more effective for clinics that were poor performers at baseline. Finally, in models that compared improvement in the population of focus (applicable for 11 intervention clinics) to control clinics, only the increase in the number of visits in 3 or 4 quarters was significant ($P = 0.04$). In addition, the results did not change when we used fixed-effect models or when we repeated the analyses for highly active antiretroviral therapy use and viral load control based on the new guidelines for antiretroviral use released in 2002.

DISCUSSION

To our knowledge, this study is the first national controlled evaluation of a Breakthrough Series collaborative.

In a prospective matched trial including almost 10 000 patients in 69 clinics, we found that the quality improvement collaborative did not significantly affect virologic outcomes or process measures, such as screening, prophylaxis, and access to care.

Quality improvement methods of various types have been widely adopted throughout the health care system, and reports of successful interventions are numerous (17, 22, 23). However, few rigorously controlled trials have examined the implementation of such programs (14, 23–29). Furthermore, publication bias favors the dissemination of successful interventions (30, 31). Thus, the evidence on the success of such efforts is mixed, and quality has generally not dramatically improved (23).

Multi-institution collaborative quality improvement programs such as the Breakthrough Series are particularly important because of their well-established and documented methods, broad applicability, and potential for widespread adoption (18). Notable examples of quality improvement collaboratives have demonstrated improvements in small numbers of sites (14, 32), and other anecdotal or less rigorously evaluated examples have shown dramatic improvements (15). These studies suggest that selected institutions can achieve substantial improvements in quality in collaboratives. Many previous studies of collaboratives, however, have had selection bias, have relied on pre–post designs without being able to account for secular trends, have used self-report measures rather than medical record reviews, or have included only selected samples from self-selected sites (14, 15, 19, 23, 32, 33). For instance, Horbar and colleagues (14) evaluated quality improvement in a collaborative that focused on neonatal care. They demonstrated improvements relative to a set of control clinics, but the 2 intervention groups were small (6 and 4 sites) and self-selected. Other evaluations, such as one of a collaborative to reduce cesarean section rates, have used historical controls that make accounting for secular trends difficult; also, these evaluations have usually not incorporated a control group (15, 19, 34). Thus, we believe that our evaluation may be the most rigorous assessment of a collaborative to date. To illustrate the importance of this, if we had conducted an uncontrolled study, we might have attributed the 11–percentage point improvement in viral load control and the 7.3–percentage point increase in the percentage of patients receiving an influenza shot to the intervention, whereas our trial showed that these increases did not differ significantly from those observed in the control clinics.

Our study has several limitations. First, we could not perform a pure randomized trial of the intervention. Instead, we relied on rigorous matching and statistical models to further adjust for potential confounding variables. Lack of randomization, however, is of less concern when examined in the context of our results. Typically, a major concern of this type of evaluation is that clinics that are more amenable to quality improvement or more familiar with

the concepts of continuous improvement would be more likely to participate in this type of intervention. Consequently, one would expect that the resulting bias would be in the direction of finding positive results. Second, although we assessed important markers of HIV quality that were the main focus of the collaborative, some clinics might have improved in areas of care that we did not measure, such as adherence. If that were the case, however, we would expect to see improvements in these areas reflected in some of the other important measures tracked by each clinic in the collaborative, such as the proportion of patients with a controlled viral load. Third, Hawthorne effects could have led to improvements in the control clinics by virtue of their being observed. Fourth, the Health Resources and Services Administration mandated participation for some clinics, and the collaborative was larger than most previous collaboratives. Several other collaboratives, however, have used similar approaches, and our subgroup analysis did not show any differences between mandated and voluntary clinics. Finally, some of the control clinics might have been “contaminated” by participation in earlier quality improvement collaboratives aimed at different clinical conditions. Although we cannot say whether previous participation might have improved performance in these clinics, no data currently suggest that organizations have been able to spread these methods throughout the organization or to other clinical conditions.

For some of the measures that we studied, such as the use of highly active antiretroviral therapy, quality of care was excellent in both intervention and control clinics, thus leaving little room for improvement. Of note, our results demonstrated that both intervention and control patients were receiving highly active antiretroviral therapy less frequently in the second period than the first. This is probably due to secular trends in the use of highly active antiretroviral therapy (20, 21). Other measures, however, such as the screening, prophylaxis, and access to care measures, also did not show any improvement, and performance on these measures was not as good. All of these process measures are widely accepted as beneficial, and their implementation is not complex. In addition, we expected that these types of measures would be good candidates for improvement because parts of the quality improvement intervention focus on the use of patient registries and other automated processes. These resources can be used to identify and then provide intervention for patients who need particular services, such as screening examinations or immunizations. It is still possible, however, that clinically important differences might have been detected with a larger sample size. The difference approached statistical significance for one of the measures ($P = 0.06$ for Papanicolaou smears).

Our study, like other evaluations of quality improvement programs, is generalizable only to the disease we studied, a chronic medical condition. Different clinical areas might be more amenable to the improvement methods

we studied. For instance, other collaboratives have focused on a particular procedure or process of care, such as rates of cesarean sections or successful extubations, which are discrete processes that might be more amenable to collaborative improvement processes. Improving care for patients with chronic medical conditions, however, has become a major focus of collaborative improvement programs (35). Collaboratives are being conducted for such conditions as asthma, diabetes, and depression. Although not generalizable to all chronic medical conditions, to our knowledge, this is the first rigorously controlled trial that examines the effectiveness of this well-known type of collaborative for a chronic medical condition. Nonetheless, because each targeted disease and each breakthrough program have idiosyncratic components, we would not conclude that all Breakthrough Series collaboratives for chronic medical conditions are ineffective. Our results, however, should give pause to those considering implementing similar collaboratives, and, if additional evidence from similar studies supports our findings, a different approach should be advocated. Alternatively, in the spirit of continuous improvement, similar results from additional evaluations might prompt a redesign of the Breakthrough Series process to potentially improve the effectiveness of the intervention.

A strength of our study is that, unlike most published studies of quality improvement interventions, we used statistical techniques that account for clustering of patients within clinics. Accounting for such clustering is even more important when the intervention is at the level of the institution as opposed to individual patients (36, 37). We also conducted several post hoc analyses to test the robustness of our results. In addition, our study was powered to detect clinically significant differences even before expanding the sample of clinics. Given the nature of quality improvement collaboratives, it is unlikely that future studies will have this many participants.

Finally, we do not know whether the collaborative failed because participants were not engaged actively enough, because of the types of changes that were implemented, or because of the way that they were implemented. Our impressions from attending the learning sessions and postintervention site visits are that most participants were actively engaged in the improvement processes. Some initiatives may have been more effective at some sites than at others and that different sites may have improved in different areas. However, if that were the case, the fact that there was not significant improvement for all the clinics means that some clinics would have had to experience declines in quality in order to balance out any selective positive effects. Future research should address this issue in more detail.

In conclusion, in this national controlled trial of a quality improvement collaborative aimed at improving care for patients with HIV infection, rates of improvement did not significantly differ between participating and control sites. Our findings suggest that additional research is

needed to improve methods of teaching and implementing quality improvement programs to achieve better results.

From Harvard Medical School, Beth Israel Deaconess Medical Center, New England Medical Center, and Dimock Community Health Center, Boston, Massachusetts; Harvard University, Cambridge, Massachusetts; and The University of Wisconsin, Madison, Wisconsin.

Acknowledgments: The authors thank their colleagues at the HIV AIDS Bureau of the Health Resources and Services Administration for their efforts and expertise. They also thank Lin Ding, PhD, for assistance with expert statistical programming; Deborah Collins for assistance with manuscript preparation; Carol Cosenza, MSW, and Patricia Gallagher, PhD, of the Center for Survey Research for assistance with instrument development and survey administration; and Barbara J. McNeil, MD, PhD, for comments on an earlier version of this manuscript.

Grant Support: By a grant from the Agency for Healthcare Research and Quality (R-01HS10227).

Potential Financial Conflicts of Interest: *Other:* D. Gustafson (board member of the Institute for Healthcare Improvement).

Requests for Single Reprints: Bruce E. Landon, MD, MBA, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115; e-mail, landon@hcp.med.harvard.edu.

Current author addresses are available at www.annals.org.

References

- Kohn LT, Corrigan JM, Donaldson MS, eds. To Err Is Human: Building a Safer Health System. Washington, DC: National Academy Press; 2000.
- McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. *N Engl J Med*. 2003;348:2635-45. [PMID: 12826639]
- Paella FJ Jr, Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA, et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *N Engl J Med*. 1998;338:853-60. [PMID: 9516219]
- Stein MD, Piette J, Mor V, Wachtel TJ, Fleishman J, Mayer KH, et al. Differences in access to zidovudine (AZT) among symptomatic HIV-infected persons. *J Gen Intern Med*. 1991;6:35-40. [PMID: 1999744]
- Shapiro MF, Morton SC, McCaffrey DF, Senterfitt JW, Fleishman JA, Perlman JF, et al. Variations in the care of HIV-infected adults in the United States: results from the HIV Cost and Services Utilization Study. *JAMA*. 1999;281:2305-15. [PMID: 10386555]
- Moore RD, Stanton D, Gopalan R, Chaisson RE. Racial differences in the use of drug therapy for HIV disease in an urban community. *N Engl J Med*. 1994;330:763-8. [PMID: 8107743]
- Shortell SM, O'Brien JL, Carman JM, Foster RW, Hughes EF, Boerstler H, et al. Assessing the impact of continuous quality improvement/total quality management: concept versus implementation. *Health Serv Res*. 1995;30:377-401. [PMID: 7782222]
- Chassin MR, Galvin RW. The urgent need to improve health care quality. Institute of Medicine National Roundtable on Health Care Quality. *JAMA*. 1998;280:1000-5. [PMID: 9749483]
- Berwick DM. Continuous improvement as an ideal in health care. *N Engl J Med*. 1989;320:53-6. [PMID: 2909878]
- Institute for Healthcare Improvement Collaboratives. Accessed at www.ihi.org/collaboratives/ on 12 Dec 2003.
- Berwick DM, Nolan TW. Physicians as leaders in improving health care: a new series in *Annals of Internal Medicine*. *Ann Intern Med*. 1998;128:289-92. [PMID: 9471932]

12. Demming EW. *Out of the Crisis*. Cambridge, MA: MIT Pr; 2000.
13. Langley GJ, Nolan KM, Norman CL, Provost LP, Nolan TW. *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance*. San Francisco: Jossey Bass; 1997.
14. Horbar JD, Rogowski J, Plsek PE, Delmore P, Edwards WH, Hocker J, et al. Collaborative quality improvement for neonatal intensive care. NIC/Q Project Investigators of the Vermont Oxford Network. *Pediatrics*. 2001;107:14-22. [PMID: 11134428]
15. Flamm BL, Berwick DM, Kabcenell A. Reducing cesarean section rates safely: lessons from a "breakthrough series" collaborative. *Birth*. 1998;25:117-24. [PMID: 9668746]
16. 2001 Ryan White CARE Act Title III: Early Intervention Services. Accessed at <http://hab.hrsa.gov/reports/tideiii2001/tideiii2001.htm> on 1 March 2003.
17. Kilo CM. A framework for collaborative improvement: lessons from the Institute for Healthcare Improvement's Breakthrough Series. *Qual Manag Health Care*. 1998;6:1-13. [PMID: 10339040]
18. Wilson T, Berwick DM, Cleary PD. What do collaborative improvement projects do? Experience from seven countries. *Jt Comm J Qual Saf*. 2003;29:85-93. [PMID: 12616923]
19. Leape LL, Kabcenell AI, Gandhi TK, Carver P, Nolan TW, Berwick DM. Reducing adverse drug events: lessons from a breakthrough series collaborative. *Jt Comm J Qual Improv*. 2000;26:321-31. [PMID: 10840664]
20. Carpenter CC, Cooper DA, Fischl MA, Gatell JM, Gazzard BG, Hammer SM, et al. Antiretroviral therapy in adults: updated recommendations of the International AIDS Society-USA Panel. *JAMA*. 2000;283:381-90. [PMID: 10647802]
21. Yeni PG, Hammer SM, Carpenter CC, Cooper DA, Fischl MA, Gatell JM, et al. Antiretroviral treatment for adult HIV infection in 2002: updated recommendations of the International AIDS Society-USA Panel. *JAMA*. 2002;288:222-35. [PMID: 12095387]
22. Blumenthal D, Kilo CM. A report card on continuous quality improvement. *Milbank Q*. 1998;76:625-48, 511. [PMID: 9879305]
23. Shortell SM, Bennett CL, Byck GR. Assessing the impact of continuous quality improvement on clinical practice: what it will take to accelerate progress. *Milbank Q*. 1998;76:593-624, 510. [PMID: 9879304]
24. Mehta RH, Montoye CK, Gallogly M, Baker P, Blount A, Faul J, et al. Improving quality of care for acute myocardial infarction: The Guidelines Applied in Practice (GAP) Initiative. *JAMA*. 2002;287:1269-76. [PMID: 11886318]
25. Holman WL, Allman RM, Sansom M, Kiefe CI, Peterson ED, Anstrom KJ, et al. Alabama coronary artery bypass grafting project: results of a statewide quality improvement initiative. *JAMA*. 2001;285:3003-10. [PMID: 11410099]
26. Wells KB, Sherbourne C, Schoenbaum M, Duan N, Meredith L, Unutzer J, et al. Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. *JAMA*. 2000;283:212-20. [PMID: 10634337]
27. O'Connor GT, Plume SK, Olmstead EM, Morton JR, Maloney CT, Nugent WC, et al. A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. The Northern New England Cardiovascular Disease Study Group. *JAMA*. 1996;275:841-6. [PMID: 8596221]
28. Soumerai SB, McLaughlin TJ, Gurwitz JH, Guadagnoli E, Hauptman PJ, Borbas C, et al. Effect of local medical opinion leaders on quality of care for acute myocardial infarction: a randomized controlled trial. *JAMA*. 1998;279:1358-63. [PMID: 9582043]
29. Shafer MA, Tebb KP, Pantell RH, Wibbelsman CJ, Neuhaus JM, Tipton AC, et al. Effect of a clinical practice improvement intervention on Chlamydia screening among adolescent girls. *JAMA*. 2002;288:2846-52. [PMID: 12472326]
30. Agema WR, Jukema JW, Zwinderman AH, van der Wall EE. A meta-analysis of the angiotensin-converting enzyme gene polymorphism and restenosis after percutaneous transluminal coronary revascularization: evidence for publication bias. *Am Heart J*. 2002;144:760-8. [PMID: 12422143]
31. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ*. 1997;315:640-5. [PMID: 9310565]
32. Pronovost PJ, Jenckes M, To M, Dorman T, Lipsett PA, Berenholtz S, et al. Reducing failed extubations in the intensive care unit. *Jt Comm J Qual Improv*. 2002;28:595-604. [PMID: 12425254]
33. Kerr D, Bevan H, Gowland B, Penny J, Berwick D. Redesigning cancer care. *BMJ*. 2002;324:164-6. [PMID: 11799037]
34. Bate P, Robert G, McLeod H. Report on the "Breakthrough" collaborative approach to quality and service improvement within four regions of the National Health Service. A research based investigation of the orthopaedic services collaborative within the Eastern, South and West and Trent regions. Research report no. 42. Birmingham, United Kingdom: University of Birmingham; 2002.
35. Health Disparities Collaboratives. Accessed at www.healthdisparities.net on 12 December 2003.
36. Divine GW, Brown JT, Frazier LM. The unit of analysis error in studies about physicians' patient care behavior. *J Gen Intern Med*. 1992;7:623-9. [PMID: 1453246]
37. Whiting-O'Keefe QE, Henke C, Simborg DW. Choosing the correct unit of analysis in Medical Care experiments. *Med Care*. 1984;22:1101-14. [PMID: 6513619]

Current Author Addresses: Drs. Landon, McInnes, Landrum, and Cleary: Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115.
Dr. Wilson: The Health Institute, Division of Clinical Care Research, Tufts–New England Medical Center, 750 Washington Street, PO Box 345, Boston, MA 02111.

Dr. Hirschhorn: 26 Berkshire Road, Newton, MA 02460.
Dr. Marsden: William James Hall, #630, Harvard University, 33 Kirkland Street, Cambridge, MA 02138.
Dr. Gustafson: Center for Health Systems Research and Analysis, WARF Building, 11th Floor, 610 Walnut Street, Madison, WI 53705-2397.