Process and Outcome: Lack of Correlation in a Primary Care Model

John C. Dickinson, MD, and Stephen H. Gehlbach, MD
Durham, North Carolina

Records of 87 hypertensive patients in a group family practice were reviewed to assess the validity of the process audit in determining medical care quality. Although a significant reduction in mean diastolic blood pressures was found at two follow-up intervals, physician performance scores showed no significant correlation with this outcome. Nor could an association between medical process and outcome be demonstrated when control of blood pressure to less than 95 mmHg was used as an outcome measure. The reliability between two chart auditors was poor, with complete agreement being achieved in only 29 percent of cases. Accurate quality assessment will require more practical methods of review which relate directly to patient outcomes.

The chart audit has become a popular tool for assessing the quality of medical practice. A number of reports describe sets of criteria which may be used as explicit standards for chart review to measure physician performance. Kessner,[1] for example, has outlined protocols for evaluating urinary tract infection, diabetes, hypertension, and iron deficiency anemia. Others have assessed tonsillitis, asthma,[2] appendicitis, myocardial infarction,[3] and peptic ulcer disease.[4] The assumption underlying these "process" audits of health care is that physicians who demonstrate high adherence to established criteria, eg, performing fundoscopic examinations, obtaining urine cultures, ordering antibiotics, are providing good care. Generally, standards are determined by panels of physicians so that process audits may be reasonable estimates of whether a physician's performance is acceptable in the eyes of his/her peers.

However, there is a growing concern that adherence to process criteria may not be the best measure of quality of care. All that physicians do to and for patients does not necessarily improve the patients' well-being. Favorable patient outcome is probably a more important index of successful medical care, and the question remains: Does correct process necessarily result in good patient outcome?

To explore the validity of the process audit the authors conducted a chart review of hypertensive patients in a family practice. If process criteria are valid measures of quality, outcome should improve as adherence to process increases.

Materials and Methods

The study population was the registered patient population of the Duke-Watts Family Medicine Residency Program. This group practice serves a population of approximately 8,000, with patients ranging in age from newborns to nonagenarians, and provides a full spectrum of medical care. Patient diagnoses are recorded by the primary provider at each patient visit, coded using the International Classification of Health Problems in Primary Care (ICHPPC),[5] and stored in a computer file.

The study group was derived from a computer-generated list of all patients with the diagnosis of high blood pressure or hypertension. Thus, false negatives (not identified by the provider) were excluded and false positives (erroneous diagnoses)

From the Duke University Medical Center, Duke-Watts Family Medicine Program, Durham, North Carolina. Requests for reprints should be addressed to Dr. Stephen H. Gehlbach, Duke-Watts Family Medicine Program, 1012 Broad Street, Durham, NC 27705.

0094-3509/78/0901-0557$01.50

Table 1. Demographic Characteristics of Total Family Medicine Center Patients Compared with Study Population

Characteristic		FMC Population N=8,160	Study Group N=87
Age	Median (range) in years	29 (0-99)	61 (21-86)
Sex	Male	2,940 (36%)	25 (29%)
	Female	5,105 (64%)	62 (71%
Race	White	6,435 (79%)	62 (71%)
	Black	1,435 (18%)	25 (29%)
	Other	298 (3%)	0 (0%)

were included. A sample of patient records was randomly selected from this list. The following variables were then defined:

1. The "index visit" was the encounter in which the diagnosis of hypertension was first made. The blood pressure recorded for this visit was averaged with a second untreated pressure to give an "index blood pressure."
2. A "first follow-up blood pressure," designed to assess short-term outcome, was the average of readings taken at two visits, 5 to 13 months after the index visit.
3. The "final follow-up pressure," assessing longer term outcome, was an average of readings taken at the patient's latest two visits.

Diastolic pressure was measured by most providers surveyed at the fifth Korotkoff sound. When both fourth and fifth sounds were recorded the latter was included for purposes of this study. For a single encounter the highest diastolic reading was used.

Two physicians independently reviewed the final group of patient records. Percentage scores for physician performance were derived from the application of a modified Kessner protocol for hypertension audit.[6] Twenty-six process criteria were given equal weight. They included 12 historical, 7 physical examination, and 5 laboratory items, as well as one each for adequate diagnosis and management. Scoring was liberal. For example, if the criterion called for examination of the jugular veins, the provider needed only indicate "neck" on an examination checklist to obtain credit. Audits were carried out by each reviewer independently. Differences in process scores between auditors were resolved by direct negotiation.

Results

Seven hundred and ninety-nine patients with a recorded diagnosis of elevated blood pressure or hypertension were identified, representing 12.5 percent of the population over 19 years of age. Of these, 212 records were "randomly" selected for review. Nine could not be located after two attempts. Six records lacked any reference to elevated blood pressure or hypertension; ie, the computer list was in error. Of the remaining 197, 110 were eliminated because of inadequate outcome data. In most cases these patients had not been enrolled in the practice long enough to qualify. The final study group included 87 patients. The demographic characteristics of this population are compared to those of the total practice in Table 1.

Of the 50 providers in the practice, 41 were represented by at least one patient in the study group. The physician performance or process scores are presented in Table 2. The average adherence score was 59 percent of the 26 minimal care criteria, with a range of 30 to 82 percent. On the average 43 percent of historical, 77 percent of physical examination, and 60 percent of laboratory items were included. The diagnosis was judged appropriate in 98 percent of records and the management in 82 percent. Fourteen of the 16 records that failed the management criterion were of patients with borderline hypertension which was untreated after five months or for whom treatment was discontinued. Another patient was started on an unacceptable medication; the last was treated but did not satisfy the protocol definition for hypertension.

The average index visit diastolic pressure was 99.6 mmHg with a range of 60 to 136 mmHg. Six-

teen patients had an index diastolic pressure less than 90 mmHg; 15 of these were already taking hypertensive medications upon entering the practice.

Eighty records contained outcome data for the first follow-up interval. The average diastolic pressure was 92.5 mmHg (range 55 to 130 mmHg). Seventy-five records qualified in the second follow-up interval and had an average diastolic reading of 82.6 mmHg (range 58 to 123 mmHg). The average fall from index to follow-up diastolic pressure was 7.1 and 17.0 mmHg in the first and second intervals, respectively. These improvements were statistically significant ($P<.05$), using Student's test for the difference between means. That is to say, it is 95 percent certain that the differences found did not occur by chance. Decreases were also seen for the subgroup of patients for whom the diagnosis of hypertension was first made in this practice (Figure 1).

Attempts to demonstrate a linear association between adherence to process criteria and reduction in blood pressure were unsuccessful. Pearson correlation coefficients relating total adherence scores to the change in blood pressure for the two follow-up periods showed no correlation ($r=0.008$ and 0.03, $P=0.47$ and 0.41). Figure 2 depicts the lack of correlation between adherence scores and

Table 2. Percent Adherence to Hypertension Process Criteria for 87 Medical Records

	Percent Adherence	
History		
1. Personal and Social History	82	
2. Family History (High Blood Pressure, Heart Disease, Stroke)	95	
3. Past Diagnosis High Blood Pressure	99	
4. Past Treatment High Blood Pressure	90	
5. History of Renal Disease	15	
6. History of Urinary Tract Infection	14	
7. History of Intravenous Pyelogram	2	
8. Chest Pain (Described)	31	
9. Ankle Swelling	17	
10. Orthopnea	20	
11. Nocturnal Dyspnea	16	
12. Shortness of Breath	26	
Average Total History Score	**43%**	**(Range 17% to 75%)**
Physical Examination		
1. Weight (½), Height (½)	25 scored ½, 75 scored 1	
2. Blood Pressure, Supine (½), Upright (½)	78 scored ½, 22 scored 1	
3. Fundoscopic	86	
4. Cardiac	91	
5. Neck (with jugular veins)	67	
6. Abdominal Examination (with bruits)	72	
7. Extremities (with pulses and edema)	75	
Average Total Examination Score	**77%**	**(Range 14% to 100%)**
Laboratory		
1. Urinalysis	86	
2. Hematocrit/Hemoglobin	72	
3. Blood Urea Nitrogen or Creatinine	72	
4. Electrocardiogram (Diastolic>130)	55	
5. Intravenous Pyelogram	9	
Average Total Laboratory Score	**60%**	**(Range 0% to 100%)**
Diagnosis	98	
Management	82	
Average Total Score	**59%**	**(Range 30% to 82%)**

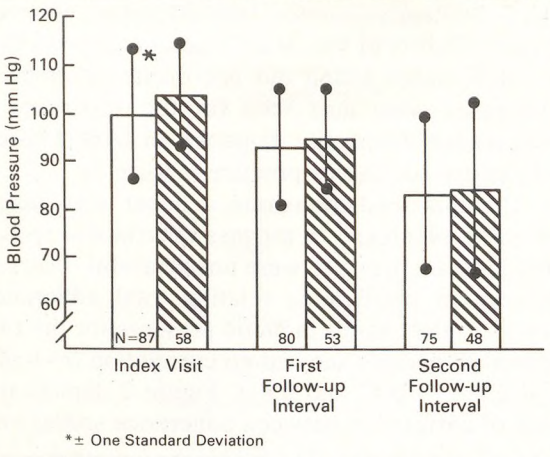

Figure 1. Mean outcome diastolic pressures for total and previously untreated populations. Clear bars are for total population, hatched bars represent sub-group of patients first diagnosed at the Duke-Watts Family Medical Center.

percent change in blood pressure in the final follow-up period. Examining data for the group of newly diagnosed patients also failed to show a significant correlation, as did analyzing by history, physical examination, and laboratory subscores of the process criteria.

The authors also compared process scores with blood pressure control. "Control" was defined as a diastolic pressure less than 95 mmHg, and "noncontrol" as 95 mmHg or above. Process scores were analyzed against these dichotomous variables using a point biserial correlation technique.[7] Again, no significant correlation was found for the final follow-up period for either the total study group or the newly diagnosed patients ($r=.058$ and $P=13$).

Over 55 physician hours were consumed in the implementation of the audit, exclusive of data analysis time, for an average of 19 minutes per chart per auditor. Interrater reliability is shown in Figure 3. Although there was obvious correlation between ratings, complete agreement was achieved for only 29 percent of the 87 records. Disagreements were noted in an average of 1.7 items per chart with a range of zero to six items.

Discussion

In the period of study, this cohort of hypertensive patients showed progressive improvement. Not only did an increasing percentage of the population achieve control, but the average fall in diastolic pressure increased over time.

A likely explanation for clinical success is that some element in the encounter between physician and patient has brought about pressure reductions. However, several alternative explanations are possible. One could argue that the requirement for specific recorded follow-up data for the study group selected against noncompliant or severely ill patients, who would be expected to do poorly. However, most of the 110 ineligible records were eliminated because they belonged to patients who were diagnosed so recently that insufficient time had elapsed to meet the follow-up criteria, not because patients were lost to follow-up. Another explanation for the improvement in blood pressures might be observer bias. That is, providers may underestimate follow-up pressures when they expect a response to treatment. It is also possible that reduced blood pressures simply reflect the effect of "regression toward the mean."[8] When measurements are made on a population of patients and only extremes of the distribution are selected for further study (ie, blood pressure>90 mmHg), the repeated measurements will tend to fall toward the mean of the parent population. Highly variable measurements, like blood pressure, show larger regression effects.

Despite the overall improvement in blood pressures, physician adherence scores were not particularly high. On the average, only 59 percent of criteria were fulfilled overall and only 43 percent of history items. Furthermore, the authors were unable to demonstrate that successful blood pressure reduction was related to the degree of adherence to the protocol. No correlation could be found even for the cohort of patients who were diagnosed for the first time at the index visit and should have shown most improvement. These conclusions corroborate recent work of Nobrega,[9] who showed that of 89 process criteria studied, only three (weight, age, and initial diastolic blood pressures) were predictive of short-term outcome.

The absence of association between recorded process and adequate outcome raises serious questions about the validity of the standardized process audit as a predictor of patient improvement. The problem may lie in the discrepancy between recorded and actual process. Physicians may comply with criteria but fail to record their

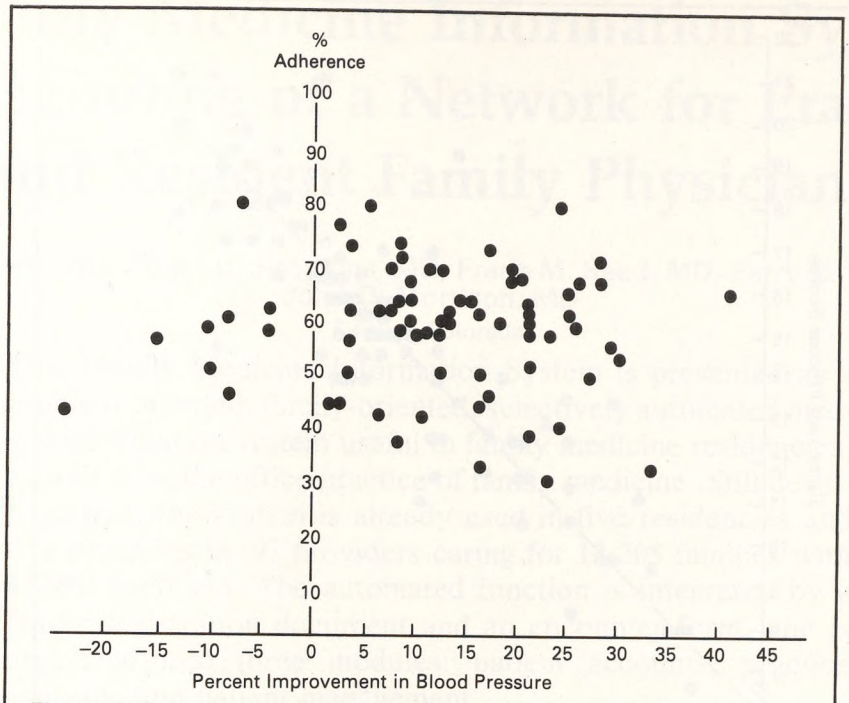

Figure 2. Percent improvement in blood pressure from index visit to latest follow-up compared to process criteria adherence scores (%) for 75 hypertensive patients.

adherence in the chart. Recent work on the relationship between recorded and actual performance supports this explanation.[10,11]

In addition, the specific items included in the usual process score may be irrelevant to outcomes. Most are derived through "expert consensus" which places a high value on differential diagnosis, less on proper management, and even less on issues of compliance and patient satisfaction. Physicians are taught to ask about a family history of hypertension and to examine a patient's fundi. But if the patient has a blood pressure of 160/110 mmHg, it does not really matter whether his grandfather had high blood pressure or if he has vascular changes in his retina. Being prescribed proper treatment and taking his medication are probably all that are essential to controlling his pressure and improving his ultimate outcome.

Items in a process criteria list may fail to correlate with disease outcome because they are designed to detect either rare etiologies for the disease (history of renal disease, abdominal bruits) or its late stage complications (ankle swelling, orthopnea, neck vein distention). Since a rather small percentage of primary practice patients will have either secondary causes for their hypertension or serious complications, a great many charts must be audited before benefits of adhering to these detailed criteria are detected. Thus, the inability to demonstrate a correlation between adherence to process and control of blood pressure in a small series of family medicine patients may be unfair. Perhaps these criteria are valid for large populations or would correlate better with outcomes that accounted for symptoms of heart failure or renal insufficiency. Unfortunately, these broader outcome validations have yet to be accomplished.

Problems with reliability were another disconcerting finding in this study. While other authors have found 90 percent reliability when results of lay abstractors were compared,[12] the two of us agreed only 29 percent of the time. While most of the disagreements could be easily resolved, we were struck with the number of judgments that were necessary in scoring even these "explicit" criteria.

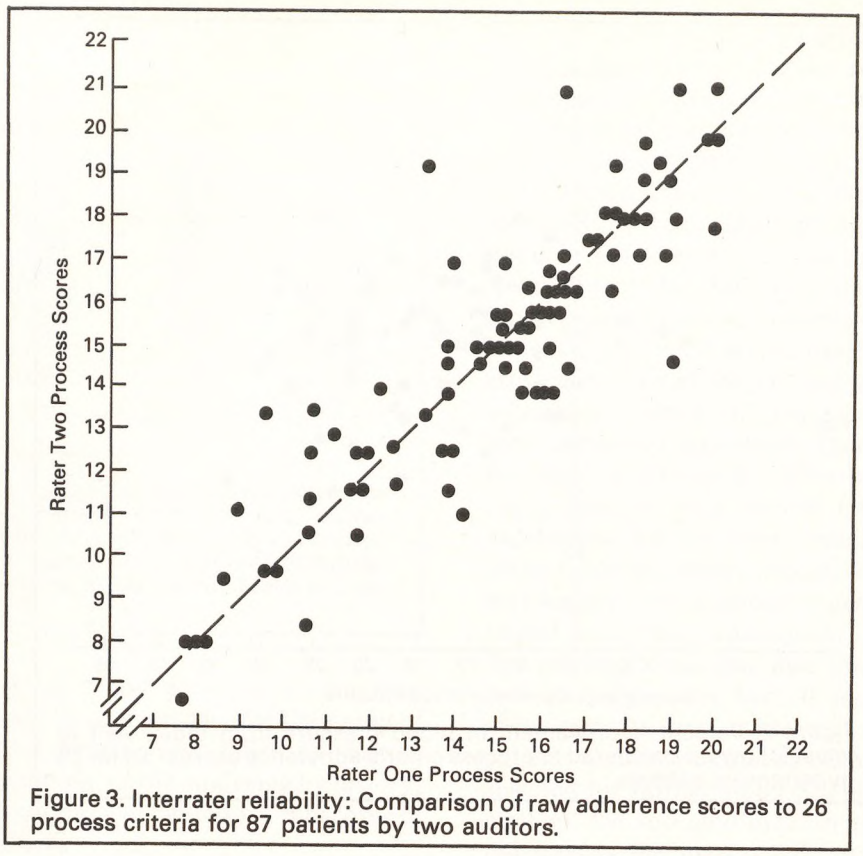

Figure 3. Interrater reliability: Comparison of raw adherence scores to 26 process criteria for 87 patients by two auditors.

The authors' reservations about process audit are shared by others. Brook et al[13] express many of the same concerns about inadequate records, lengthy audit times, and lack of process validity. In a detailed review of health status indicators, they suggest methods for developing disease-specific outcome standards. However, at present, practical instruments suitable for use in practices are not available. Better measurements must be devised that truly reflect the patient's response to medical care; they must be easily applied so that repeated audits may be accomplished, and they must be applicable to a wide variety of clinical settings.

Acknowledgement

This project was supported in part by funds from the Robert Wood Johnson Foundation.

References

1. Kessner DM, Kalk CE, Singer J: Assessing health quality: The case for tracers. N Engl J Med 288:189, 1973

2. Thompson HC, Osborne CE: Quality assurance of ambulatory child health care: Opinions of practicing physicians about proposed criteria. Med Care 14:22, 1976

3. Fessel WJ, VanBrunt EE: Assessing quality of care from the medical record. N Engl J Med 286:134, 1972

4. Brook RH, Appel FA: Quality-of-care assessment: Choosing a method for peer review. N Engl J Med 288:1323, 1973

5. International classification of health problems in primary care. Report of the Classification Committee of the World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians. Chicago, American Hospital Association, 1975

6. Smith SR: Application of the tracer technique in studying quality of care. J Fam Pract 1(3/4):38, 1974

7. Ferguson GA: Statistical Analysis in Psychology and Education. New York, McGraw-Hill, 1966

8. Isaac S, Michael WB: Handbook in Research and Evaluation. San Diego, Calif, Edits, 1971

9. Nobrega FT, Morrow GW Jr, Smoldt RK, et al: Quality assessment in hypertension: Analysis of process and outcome methods. N Engl J Med 296:145, 1977

10. Tufo HM, Speidel JJ: Problems with medical records. Med Care 9:509, 1971

11. Bentsen BG: The accuracy of recording patient problems in family practice. J Med Educ 51:311, 1976

12. Payne BC: The Duality of Medical Care: Evaluation and Development. Chicago, Hospital Research and Evaluation Trust, 1976

13. Brook RH, Davis-Avery A, Greenfield S, et al: Assessing the quality of medical care using outcome measures: An overview of the method. Med Care 15(suppl to No. 9):1-165, 1977