Characterizing Patient Requests and Physician Responses in Office Practice

Richard L. Kravitz, Robert A. Bell, Carol E. Franz, Marc N. Elliott, Ezra Amsterdam, Carrie Willis, and Lisa Silverio

Objective. To assess the reliability, applicability, and validity of a refined system (taxonomy of requests by patients [TORP]) for characterizing patient requests and physician responses in office practice.

Study Settings. Data were obtained from visits to six general internists practicing in North-Central California in 1994 and eight cardiologists practicing in the same region in 1998.

Study Design. This was an observational study of patient requests and physician responses in two practice settings. Patients were surveyed before and after the visit. Physicians were surveyed immediately after the visit, and all visits were audio recorded for future study.

Data Collection/Extraction Methods. TORP was refined using input from a multidisciplinary panel. Audiotape recordings of 131 visits (71 in internal medicine and 60 in cardiology) were rated independently by two coders. Estimates of classifying reliability (intercoder agreement on the sorting of requests into categories) and unitizing reliability (intercoder agreement on the labeling of elements of discourse as "requests" and subsequent classification into categories) were calculated. Validity was assessed by testing three specific hypotheses concerning the antecedents and consequences of patient requests and request fulfillment.

Principal Findings. The overall unitizing kappa for identifying patients' requests was 0.64, and the classification kappa was 0.73, indicating substantial agreement beyond chance. The average patient made 4.19 requests for information and 0.88 requests for physician action; there were few differences in the spectrum of requests between internal medicine and cardiology. Approximately 15 percent of visits included a direct request for completion of paperwork. Patients who were very or extremely worried about their health made more requests than those who were not (6.06 vs. 3.89, $p < 0.05$). Visits involving more patient requests took longer ($p < 0.05$) and were perceived as more demanding by the treating physician ($p = 0.025$). The vast majority of requests were fulfilled.

Conclusions. The refined TORP shows evidence of both unitizing and classification reliability and should be a useful tool for understanding the clinical negotiation. In addition, the system appears applicable to both generalist and specialist practices. More experience with the system is necessary to appraise TORP's ability to predict important clinical outcomes.

More than 20 years ago, Lazare et al. conceptualized the clinical encounter as a process of negotiation between the clinician and patient (Lazare, Eisenthal, and Wasserman 1975). "Patients are conceived of as appearing with one or more requests…. It is the clinician's task to elicit the request, collect the relevant clinical data, and enter into a negotiation that should foster a relationship of mutual influence between patient and clinician" (p. 553). At the time this notion seemed radical, running counter to prevailing models of medical practice. Now the view of the clinical encounter as a two-way exchange is increasingly accepted, in part because growing evidence supports the hypothesis that increased patient involvement in the health care process produces better health outcomes (Greenfield, Kaplan, and Ware 1985). In addition, the consumer movement has so affected health care that what Lazare called the "customer approach to patienthood" has now entered the mainstream (Bartlett 1999).

Despite recognition of the importance of the clinical negotiation in office practice, the negotiation process itself has been described infrequently (Larsen, Risor, and Putnam 1997; Weijer, Singer, and Dickens 1998). However, for several reasons, efforts to understand this process have

---

This work was supported by grants from the Agency for Healthcare Policy and Research (R03 HS09812-01) and the Robert Wood Johnson Foundation (#034384). The authors thank Gary Stein, M.D., for assistance in developing the taxonomy and the participating physicians and patients for making this work possible.

Address correspondence to Richard L. Kravitz, M.D., Professor of Medicine and Director, UC Davis Center for Health Services Research in Primary Care, 4150 V. Street, Sacramento, CA 95817. Robert A. Bell, Ph.D., is a professor, UC Davis Department of Communication, Davis, CA. Carol E. Franz, Ph.D., is a research associate, UC Davis Department of Psychiatry, Sacramento, CA. Marc N. Elliott, Ph.D., is a Statistician, RAND Statistics Group, Santa Monica, CA. Ezra Amsterdam, M.D., is a professor, UC Davis Division of Cardiovascular Medicine, Sacramento, CA. Carrie Willis, B.S., is a research assistant, and Lisa Silverio, is a research assistant, UC Davis Center for Health Services Research in Primary Care, Sacramento, CA. This article, submitted to *Health Services Research* on April 19, 2000, was revised and accepted for publication on September 28, 2000.

assumed increasing urgency. Managed care has created greater opportunities for physician–patient conflict. Health care providers have increasing economic stakes in creating satisfied patients. Office visits are shorter, and physicians feel pressure to do more in less time. Physicians and patients need strategies for exchanging clinical information and reaching sound decisions efficiently in an atmosphere of mutual respect.

Requests for information or for medical services are the fulcrum for patient-initiated action in office practice, but they can sometimes be unwelcome because they consume time and resources. In particular, requests for diagnostic tests, medications, and referrals can be costly to capitated practices and may foment patient–physician discord or distrust if not handled properly. Even when requests for services are clearly not indicated, saying "no" while preserving the patient–provider relationship is challenging (Gallagher et al. 1997).

An important prerequisite for understanding the clinical negotiation is a set of tools specifically designed for studying patients' requests. The Roter Interactional Analysis System (Roter, Stewart, Punam, et al. 1997), the Davis Observation Code (Callahan and Bertakis 1991) and the Multi-Dimensional Interaction Analysis system (Charon, Greene, and Adelman 1994) are valid and reliable systems for coding physician–patient interactions, but they do not focus on the request–response process. On the other hand, the excellent systems developed by Like and Zyzanski (1986) and Eisenthal, Koopman, and Stoeckle (1990) refer to "requests" but actually rely on patient self-reports and focus on desires (patients' wishes) rather than what patients ask of their doctors.

We recently described a system for systematically classifying patients' requests and physicians' responses called TORP—a taxonomy of requests by patients; (Kravitz, Bell, and Franz 1999). Preliminary data suggest that TORP can reliably capture and categorize patients' requests in primary care. We conducted the current study to evaluate major changes in the classification scheme introduced since publication of the original article, to conduct more detailed assessment of interrater reliability, and to assess TORP's performance outside the primary care sector. In so doing, we addressed three main research questions. First, using an updated version of TORP, can coders identify and classify patients' requests in a reliable fashion? In contrast to widespread practice, interrater reliability should be evaluated as a two-stage process that consists of a unitizing component (consistency in determining whether a given element constitutes a unit of interest) and a classification component (consistency in applying category

labels). In this study, we evaluate both components of reliability. Second, how does the spectrum of patients' requests differ between general internal medicine and cardiology? Understanding our comparison may be limited because of differing data collection times (1994 vs. 1998), we chose to compare these two specialties because both physician groups care for patients with some of the same high-prevalence diseases. However, the two specialties differ in terms of patient mix, diagnostic focus, and emphasis on procedures (Kravitz, Greenfield, Rogers, et al. 1992). Thus, one would expect to see differences in the spectrum of patient requests across the two specialties. Third, to what extent does TORP demonstrate construct validity when tested in a new sample of patients under the care of internists and cardiologists? We hypothesized that patients with more health-related concerns would make more requests, greater request fulfillment would be associated with increased patient satisfaction, and visits characterized by more requests would be construed as more, demanding by physicians.

METHODS

Development and Refinement of the Taxonomy

The original TORP was assembled based on the investigators' clinical experience and the results of two patient focus groups. This taxonomy had 11 categories of patient requests for information and 8 categories of requests for action. Physician responses to patient requests were coded as one of eight mutually exclusive categories modified from Roter and Hall (1992). Reliability and validity were evaluated on 139 general internal medicine encounters obtained during a study of patients' expectations conducted in 1994 (hereafter termed the "developmental sample;" Kravitz, Callahan, Azari, et al. 1997; Kravitz, Bell, and Franz 1999).

To assure the relevance of the original taxonomy to the managed care environment of the late 1990s, we convened in mid-1998 a panel consisting of a general internist, a family physician, a cardiologist, and a research psychologist. Working independently, panel members sorted 100 index cards containing verbatim requests culled from the developmental sample into coherent groups (subcategories). The number of subcategories identified by individual panelists ranged from 8 to 12. Following group discussion, a tentative revision of the request coding was established by consensus. Request "modifiers" and new physician response codes were added (discussed later). Then, in a series of meetings extending well into

1999, the research team tested the revised scheme on approximately 20 audiotapes sampled from the original 139. As a result of this preliminary work, several codes were deleted or consolidated, and some were added. An algorithm for determining whether or not a stream of speech constitutes a "request" is provided in Figure 1. The final set of codes is presented in Table 1.

Data Collection. Data for this study were collected in two phases. As alluded to earlier, phase I (internal medicine) data were collected as part of a larger project on patients' expectations for care (Kravitz, Callahan, Azari,

Figure 1: Algorithm for determining the presence of a "request." Coders are instructed to listen to each patient utterance and address a sequence of questions. Once coders ascertain an utterance represents a request, they proceed to classify the request as 1 of 12 types of requests for information or 1 of 8 types of requests for physician action

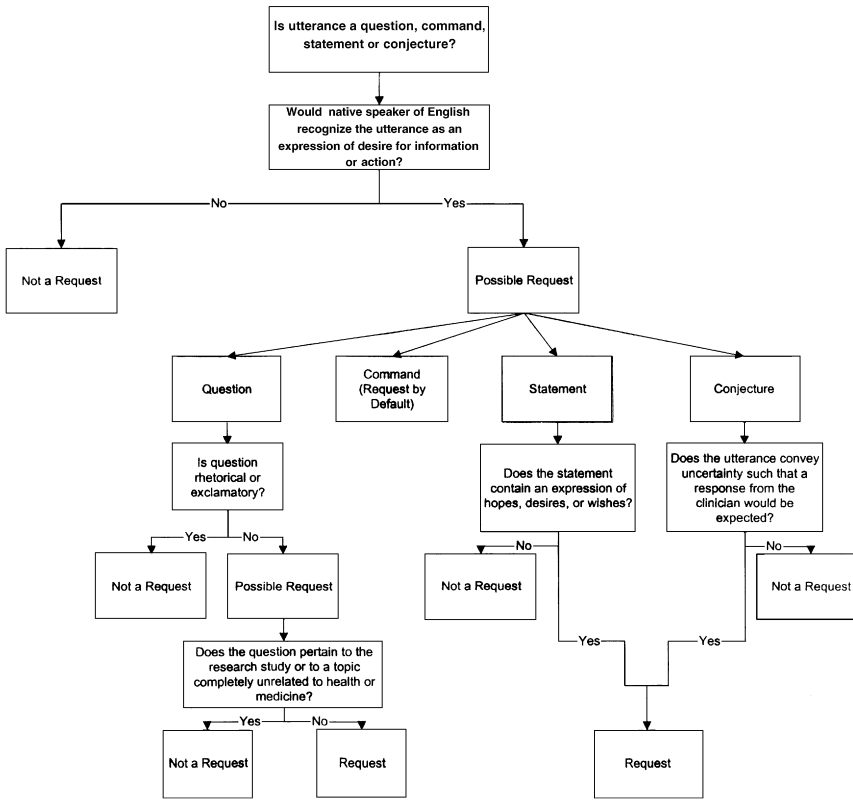

Table 1: A Taxonomy of Requests by Patients (TORP)

Requests for Information	*Requests for Action*	*Modifiers and Response Codes*
About physical symptoms, problems, or diseases	For all or part of the physical examination	Information request modifiers: occult patient, third party request, third party influence
About problems of a psychological, social, or emotional nature	For laboratory testing, imaging tests, or other diagnostic studies	Action request modifiers: occult patient, managed care, third-party request, third-party influence
About the physical examination	For a new medication or new form/dose/route of an old medication	Action request process codes: fill out form, write letter/memo, speak with third party
About diagnostic test indications, procedures, or interpretation	For refill or renewal of medication taken or prescribed in the past month	Physician response codes: acknowledgement, exploration, negotiation, fulfillment
About diagnostic test results	For a therapeutic procedure or lifestyle prescription	Levels of information request fulfillment: not fulfilled, terse, intermediate, elaborate
About drug therapy	For referral to a physician specialist	
About nondrug therapy	For referral to a nonphysician	
About preventive care	For other services	
About the current provider		
About other clinicians		
About insurance, managed care, or financial issues		
Not otherwise classifiable		

et al. 1997). Over several months in early 1994, we telephoned patients scheduled to see 1/6 general internists practicing in a single large office in a midsized northern California city. Patients were eligible for enrollment if they were at least 18 years of age and could speak and understand English. Of 503 eligible individuals, 396 (79 percent) agreed to participate, and complete data were available for 318. As described previously, we performed preliminary tests of reliability and validity on a random ("developmental") sample of 139 (Kravitz, Bell, and Franz 1999). Of the remaining 159 cases, we randomly selected 71 for detailed audiocoding in support of the current analysis.

To assess the applicability of TORP to nonprimary care settings, phase II (cardiology) data were collected over a three month period in the summer of 1998. Cardiology was selected as the target subspecialty because the spectrum of cardiology practice overlaps substantially with that of internal medicine and family medicine. Using procedures identical to those of phase I, we enrolled 60 patients visiting five cardiologists practicing within a single academically affiliated cardiology group at one of three Sacramento-area sites.

There were 131 patients in the final analytic sample for the current study (71 visiting internists during phase I data collection and 60 visiting cardiologists during phase II data collection). The median age was 55; 44 percent were women, and 25 percent were non-White. Fourteen percent did not graduate from high school. Twenty-three percent finished high school. Twenty-eight percent had at least some college, and 31 percent finished college or graduate school (5 percent declined to answer).

Questionnaire Measures. All patients completed previsit and postvisit questionnaires which included items on demographic characteristics (age, sex, ethnicity, and education), general health perceptions ("In general, would you say your health is: excellent, very good, good, fair, or poor"), health worry ("How worried are you about your health: extremely…not at all"), disease-related concerns ("How concerned are you that you might have a serious disease or condition: extremely…not at all?"), and a chronic disease count derived from a 12-item checklist completed by the treating physician. Patient satisfaction with the visit was assessed with a five-item visit-specific scale (Ware and Hays 1988). Physicians' perceptions of visit demandingness were assessed by asking, "Compared to your average patient visit, how demanding would you rate this visit in terms of the amount of effort required?" (1, far more demanding than average, to 5, far less demanding than average).

Audio-coding Procedures. Guided by a standardized training manual, trained coders listened to each tape until they came to an utterance recognized as a patient request. Coders transcribed the request verbatim (for later verification) and classified the request into one of 12 information or 8 action categories. They also marked each request with up to four possible modifiers: (1) "occult patient" (request made by patient on behalf of another person), (2) "managed care" (interaction involving managed care issue), (3) "third-party request" (third party in exam room makes request on behalf of patient), and (4) "third-party influence" (third party influences patient to make request). Although third party influences are often implicit, we instructed coders to use this modifier only when the influence was explicit and unequivocal. In addition, for action requests only, coders specified the process by which the requested action was to be accomplished, that is, by (1) completing a form, (2) writing a letter, (3) making a phone call, or (4) through unspecified means.

After classifying the request and applying appropriate modifiers and process codes, coders listened for physician responses to the patient requests. Response dimensions included (1) acknowledgment of request (yes/no), (2) exploration of request (yes/no), (3) negotiation over request (presence/absence of reciprocating exchange aimed at persuasion), and (4) resolution of request. Physician responses to requests for information were scored on a 0 to 3 categorical scale, where 0 is the patient's question dodged or ignored, 1 is a straightforward response representing only one idea, fact or opinion ("terse" response), 2 is an intermediate response supplying two distinct ideas, facts, or opinions, and 3 is a conceptually rich response supplying more than two distinct ideas, facts, or opinions ("elaborate" response). Physician responses to requests for action were coded as fulfilled, partially fulfilled, not fulfilled, or denied. All tapes were rated by two coders, with disagreements adjudicated by a senior coauthor (C.E.F.).

Assessment of Reliability. The issue of coding reliability is much more complex than has often been assumed (Wasserman and Inui 1983). Garvin, Kennedy, and Cissna (1988) distinguish unitizing reliability (consistency in identifying what is to be categorized) from classification reliability (consistency in applying category labels). Most published assessments of interrater reliability in the medical and health services literature ignore this distinction and report classification reliability only. In our own previously published analysis of TORP's interrater reliability, a trained coder identified patient requests, categorized them according to the original

TORP, and transcribed them verbatim for subsequent review by a physician-investigator. Overall agreement was 94 percent (classification kappa $= 0.93$), but this method may overestimate classification reliability (because the second rater's judgment is confirmatory and therefore not fully independent of the first rater's) and cannot evaluate unitizing reliability. In the current analysis, we assessed both unitizing and classification reliability using a fresh sample of visits.

Kappa is a chance-corrected measure of agreement. It can be computed for binary judgments (such as determining whether a particular patient utterance represents a "request") or polytomous judgments (such as classifying a request into one of several request categories). Computation of kappa involves calculation both of the actual rate of agreement ("positive agreement" and "negative agreement" for binary judgments) and the rate of agreement that would have been by chance. With binary judgments, the rate expected by chance is calculated by using (1) the proportion of all items that were marked "positive" by rater 1 (p_1) and (2) the proportion of all items that were marked "positive" by rater 2 (p_2). These proportions are then interpreted as probabilities, and the probability of agreements under the assumption of statistical independence of the raters (knowing the judgment of rater 1 for a given item supplies no information about the judgment of rater 2 for that item) is computed. The probability of positive agreement by chance is $(p_1)(p_2)$; the probability of negative agreement by chance is $(1-p_1)(1-p_2)$. The probability of any agreement by chance for a given item is therefore $(p_1)(p_2) + (1-p_1)(1-p_2)$.

The calculation of the proportions p_1 and p_2 requires identifying items that might have been rated positive by either rater. In other words, we must count the total number of items or units considered by the judges in making their ratings. For many judgments, the units are discrete and obvious, but in continuous streams of speech, this may not be so. In particular, utterances that are rated "positive" or "unitized" by at least one rater are flagged for analysis, whereas those "not unitized" or implicitly rated "negative" by both raters are not. In other words, we are concerned about the utterances that both raters agree constitute a "nonrequest" or negative agreement.

One might naively overlook and exclude utterances that neither coder recognizes as a "request," but doing so would invalidate the principles behind the kappa measure by (1) ignoring one of the sources of observed agreement and (2) making the calculation of chance rates of agreement impossible. Because the discreet boundaries of utterances not

marked positive by either rater may be difficult to discern, we estimated the total number of units (utterances) subject to judgment, that is, the total number of requests and nonrequests uttered by patients.

Our procedure was designed to generate an estimate of the number of negative agreements per minute of recorded time for each of the 131 audiotapes. We define a "negative agreement" as a discrete segment of patient speech that both raters implicitly designate as a "nonrequest." For example, consider the following transcript segment: (1) Doctor (D): So I would recommend continuing that, using the nitroglycerin. Do you have more? (2) Patient (P): I lost it and I intended to ask you to… (3) D: Yeah, we should give you some more nitroglycerin [further exchange]. (4) D: Maybe you're the type of person who can stay up a little later and still feel like you get enough sleep. (5) P: Well, not going to sleep bothers me. (6) D: The anxiety, eh? (7) P: Is there some way of taking Ativan?

In this exchange, lines 2 , 5, and 7 are segments of patient speech. Assume that two coders both judge line 2 as a request for physician action—medication refill—and line 7 as a request for physician action—new medication. The important point is that both coders implicitly judge line 5 to be a nonrequest. Our procedure (outlined later here), gives appropriate credit for these implicit agreements.

Using a sample of 13 typewritten transcripts—double coded (by L.S. and C.W.) and adjudicated (by C.E.F.), we counted within each transcript the number of lines of patient speech that were not part of a patient request. Next, we divided this figure by the average length of a patient request in that transcript (rounded up to the nearest quarter line) to yield an estimate of the number of negative agreements per transcript line. We applied the median estimate from the sample of 13 transcripts to the remaining tapes, using the average number of transcript lines per minute of audiotape to convert between lines (of transcript) and minutes (of audiotape). The median number of negative agreements per encounter estimated by this procedure was 63 (range 10 to 346).

This procedure assumes that utterances for which there was negative agreement are of the same "transcript length" as utterances that were rated positive by at least one rater. We tested the sensitivity of kappa to this assumption by recomputing kappa under the very conservative assumption that the actual number of negative agreements was 1/10th of what was estimated by the method just described. The conservatively corrected kappas differ very little from the original estimates and are therefore not reported.

Assessment of Validity and Impact. We performed three tests of validity. First, we assessed whether more requests were associated with worse health, as represented by health worry; concerns regarding the possibility of a serious disease; general health perceptions; and number of physician-reported chronic conditions. Next, we estimated the association between the proportion of requests coded as fulfilled and visit satisfaction. Finally, we assessed the impact of patient request behavior on physician perceptions of visit demandingness. To test these associations, we used correlational analysis, *t* tests, and analysis of variance as implemented in Stata 6.0 (Statacorp, 1999).

There is increasing awareness among health care researchers that standard statistical methods relying on independence between subjects may not always be appropriate for analysis of clustered data (e.g., when patients are selected from within the same physician practice; Ukoumunne, Gulliford, Chinn, et al. 1999). Correction for cluster effects tends to produce larger standard errors and wider confidence intervals but does not change the point estimates. Such correction is particularly important when making population estimates (e.g., the level of satisfaction experienced by patients in a particular community). We did not correct for clustering in our analysis because we do not purport to generalize beyond our convenience sample of 11 physicians. This study addresses methodological questions concerning the construct validity of a coding system, not conventional hypotheses about the nature of physician–patient interactions. Thus, within the 11 practices, the reported *p* values allow correct inferences about the relationships between coded requests and patient reports of health and satisfaction. These relationships form the basis for our conclusions about TORP's validity.

RESULTS

Reliability

Table 2 presents unitizing and classification kappa coefficients for each coding category individually and for all categories combined. Unitizing kappa coefficients ranged from 0.20 (request for nondrug therapy) to 0.68 (request for test results). Among the 19 request categories analyzed (2 referral categories were combined), 14 (74 percent) exhibited kappa coefficients exceeding 0.40 (consistent with moderate agreement beyond

chance). The overall unitizing kappa was 0.64, indicating substantial agreement beyond chance (Table 2).

The combined interpretive kappa coefficient was 0.73 (Table 2). Six of 19 categories had kappa statistics exceeding 0.80, representing "almost perfect" agreement beyond chance. Two categories (information: preventive care and action: nondrug treatments) displayed only fair agreement beyond chance. For coding of *physician responses* to information requests (along a 0 to 3 ordered, categorical scale), weighted agreement was 85.8 percent, expected agreement 68.3 percent, and kappa = 0.55.

Spectrum of Requests in Internal Medicine and Cardiology. For the 131 patient visits in the sample, the mean number of coded requests was 5.07 (4.19 information requests and 0.88 action requests). For presentational purposes, we coded all action requests with a "fill out form" or "write

Table 2: Unitizing and Classification Reliabilities

	Corrected Unitizing Reliability		*Classification Reliability*	
	Percent Agreement (n = 10,852)	*Kappa*	*Percent Agreement (n = 501)*	*Kappa*
Information requests				
Physical problem	99.2	0.59	92.8	0.74
Psychological problem	99.9	0.67	100	1.0
Physical examination	99.7	0.41	96.4	0.53
Diagnostic tests	99.6	0.59	96.4	0.76
Test results	99.8	0.68	97.8	0.77
Medication	99.1	0.64	95.4	0.86
Nondrug treatment	99.6	0.65	95.8	0.76
Preventive care	99.8	0.22	96.6	0.24
Index physician	99.7	0.53	98.4	0.82
Other physician	99.9	0.31	98.8	0.40
Insurance/managed care	99.9	0.45	99.0	0.66
Other information request	99.5	0.56	95.4	0.73
Action requests				
Physical examination	99.8	0.48	99.4	0.84
Diagnostic test	99.9	0.50	98.8	0.62
New medication	99.8	0.37	97.0	0.47
Medication refill	99.8	0.64	98.2	0.80
Nondrug therapy	99.9	0.20	98.8	0.24
Referral[a]	99.9	0.33	99.4	0.66
Other action request	99.6	0.42	96.8	0.62
Total	95.7	0.64	75.6	0.73

[a] Referrals to physicians and non-physicians combined for this analysis.

letter" modifier as "request for action: paperwork," regardless of the primary TORP category. There were only six visits (all in internal medicine) in which patients made no requests. The most frequent information requests concerned drug therapy (25.7 percent); physical symptoms, problems, or diseases (16.9 percent); diagnostic tests (9.5 percent); and nondrug treatments (9.5 percent). The most frequent action requests were for medication refills (27.8 percent), completion of paperwork (14.8 percent), new prescriptions (13.0 percent), and physical examination (12.2 percent).

Table 3 compares by specialty (1) the distribution of request types and (2) the proportion of encounters with at least one request of each type. The distribution of information request types was similar among patients seeing internists and cardiologists, except that cardiology patients made substantially more requests for information concerning "nondrug treatment." At least one such request was made by 12 cardiology patients compared with 5 internal medicine patients (20 vs. 7 percent, $p = 0.053$). The 12 cardiology patients made 45 such requests: as expected, a large proportion (56 percent) pertained to invasive procedures such as coronary angiography, coronary artery bypass surgery, valve surgery, and automatic defibrillator placement.

The distribution of action requests between the two specialties was similar. Only one comparison approached statistical significance: 13 percent of internal medicine patients specifically requested a component of the physical examination compared with 3 percent of cardiology patients ($p = 0.11$). Of the 12 physical examination-related requests in internal medicine, 7 were for examination of the skin or subcutaneous tissue (e.g., "please look at this mole"), 3 for examination of the extremities ("what do you think of this swelling down here?"), and 2 for examination of the ears, nose, and/or throat ("I wanted you to check my ears").

Request Modifiers and Process Codes. As previously noted, patients made at least 1 information request in 119 of the 131 visits (91 percent). Among the 119 visits, 2 included requests made on behalf of an "occult patient;" 10 included "third-party" requests (made by a visit companion), and 15 showed evidence of third-party influence (request prompted by another person or collective). Among 64 visits in which patients made at least one action request, 3 included an "occult patient" request. Five included a "managed care" request. Five had a third-party request made by a visit companion, and 13 were coded for third-party influence. The distribution of these modifier codes did not differ significantly by specialty.

Table 3: Distribution of Requests for Internal Medicine Visits ($n = 71$) and Cardiology Visits ($n = 60$)

	Internal Medicine				*Cardiology*			
	Distribution of Requests		*Visits with at least One Such Request*		*Distribution of Requests*		*Visits with at least One Such Request*	
Request Type	*No.*	*Percent*	*No.*	*Percent*	*No.*	*Percent*	*No.*	*Percent*
Information request about								
Physical symptoms, problems, disease	39	15.3	25	32.5	54	18.4	27	45.0
Psychosocial problems	4	1.6	3	4.2	2	0.7	2	3.3
Physical examination	16	6.3	10	14.1	17	5.8	13	21.7
Diagnostic test indications, procedures, or interpretation	26	10.2	14	19.7	26	8.8	18	30.0
Diagnostic test results	22	8.6	16	22.5	12	4.1	10	16.7
Drug therapy	71	27.8	30	42.3	70	23.8	37	61.7
Nondrug treatments	7	2.7	5	7.0	45	15.3	12	20.0
Preventive care	9	3.5	7	9.9	9	3.1	7	11.7
Patient–physician relationship	15	5.9	9	12.7	22	7.5	11	18.3
Other healthcare providers	3	1.2	2	2.8	4	1.4	4	6.7
Insurance, managed care, financial issues	6	2.4	3	4.2	2	0.7	2	3.3
Other requests for information	37	14.5	29	40.8	31	10.5	28	46.7
Total information requests	255	100.0			294	100.0		
Visits with ≥ one information request			60	84.5			59	98.3

Action request for								
All or part of the physical examination	12	18.2	9	12.7	2	4.1	2	3.3
Laboratory tests, imaging studies, or other diagnostic tests	5	7.6	5	7.0	2	4.1	2	3.3
New medication or new form, dose or, route of an old medication	7	10.6	6	8.5	8	16.3	7	11.7
Refill of an old medication	15	22.7	13	18.3	17	34.7	12	20.0
Nondrug treatment	3	4.5	2	2.8	1	2.0	1	1.7
Referral to another physician	7	10.6	6	8.5	2	4.1	2	3.3
Referral to a nonphysician	1	1.5	1	1.4	1	2.0	1	1.7
Completion of paperwork	10	15.1	9	12.7	9	18.4	7	11.7
Other requests for action	6	9.1	4	5.6	7	14.3	6	10.0
Total action requests	66	100.0			49	100.0		
Visits with ≥ one action request			38	53.5			26	43.3
All requests (information + action)								
Total requests			321			343		
Mean number of total requests			4.52			5.72		
(SD)			(3.99)			(4.10)		

Physician Responses to Information Requests. Physicians fulfilled almost all patient requests for information. For the 60 internal medicine patients who made at least one information request, 3 percent of the requests were not fulfilled. Thirty-two percent were fulfilled with a "terse" response; 33 percent were fulfilled with an "intermediate" response; and 32 percent with an "elaborate" response. For the 59 cardiology patients who made at least one information request, 3 percent of requests were not fulfilled, and the percentages of requests fulfilled at the "terse," "intermediate," and "elaborate" levels were 35, 28, and 34 percent, respectively.

Physician Responses to Action Requests. A total of 38 internal medicine patients and 26 cardiology patients made at least one action request. For internal medicine and cardiology patients alike, no action requests were denied outright, but 8 percent were passively unfulfilled (usually by the physician ignoring the request or changing the subject). Just over 7 percent of internal medicine patients' action requests were partially fulfilled, and 85 percent were fulfilled completely. The corresponding figures for cardiology were 16 and 76 percent, respectively.

Relationship of Requests to Visit Length and Health Status. In both specialties, the number of information requests and total requests was positively associated with visit time, as measured by tape recorder counter cycles ($p < 0.05$). The number of action requests was associated significantly with longer visits for internal medicine ($p < 0.05$) but not cardiology patients.

We expected that patients with greater health-related concerns would make more frequent requests. Patients reporting that they were very or extremely worried about their health made at least half-again as many information requests as those who were somewhat, not very, or not at all worried (mean 6.06 vs. 3.89, $p < 0.05$). For internal medicine, the correlation between health worry and information requests was $r_{pb} = 0.15$ ($p = 0.21$, $n = 69$). For cardiology patients, the correlation was $r_{pb} = 0.26$ ($p = 0.049$, $n = 60$). Similarly, patients who were very or extremely concerned about the possibility of a serious undiagnosed disease made more information requests than those who were less concerned (5.63 vs. 3.76, $p < 0.05$). For internal medicine patients, the correlation was $r = 0.02$, $p = 0.89$, $n = 67$, and for cardiology patients, $r = 0.32$, $p = 0.014$, $n = 60$. There was no significant relationship between request behavior and either general health perceptions or the number of physician-reported chronic conditions (data not shown).

Patient Satisfaction. Among patients making at least one request for physician action, the proportion of action requests fulfilled was moderately correlated with patient satisfaction (Spearman $r = 0.26$ for internal medicine and 0.43 for cardiology; $p = 0.16$ and 0.037, respectively). The proportion of information requests fulfilled and patient satisfaction was essentially uncorrelated (data not shown).

Visit Demandingness. Physicians rated visits in which patients made more information, action, and total requests as more demanding (Table 4). A strong positive association between patient requests and physician ratings of demandingness was observed for both shorter (< 11 minutes) and longer (≥ 11 minutes) visits, suggesting that request count was not simply a proxy for visit length.

DISCUSSION

The results of this study support three primary conclusions. First, trained coders can identify and classify patients' requests from audiotapes in a fairly reliable fashion. However, in our experience, the process of sorting requests into categories is substantially more reliable than the process of identifying patient requests in the first place. Second, TORP appears to be applicable to the subspecialty sector as well as primary care. The number of unclassifiable requests was similar in the two specialties. However, this conclusion must be tempered by the understanding that the data from the cardiology clinic were collected 4 years later than the data from primary care. Third, the expanded TORP demonstrates sufficient face, content, and construct validity to justify its use in future studies of the physician–patient relationship.

Although interrater reliability is commonly measured and reported in medical research publications, the distinction between unitizing and classification (interpretive) reliability is generally neglected. (In fact, a Medline search of the term "unitizing reliability" produced zero hits, whereas a PsychINFO search produced several.) In some cases, this neglect is justified because the coding unit is self-defined. For example, MacKenzie, Steinwachs, Bone, et al. (1992) studied the interrater reliability of preventable death judgments. Three panels classified trauma-related deaths as preventable, possibly preventable, or not preventable. No judgment was required in identifying the "unit" for analysis because only patients who had died from trauma were included in the sample. This

Table 4: Physicians' Ratings of Visit Demandingness as a Function of Visit Length and Number of Patient Requests

	Short Visits (< 11 minutes)		*Long Visits (≥ 11 minutes)*		*All Visits*	
	Number of Visits	*Mean Demandingness Rating (SD)*	*Number of Visits*	*Mean Demandingness Rating (SD)*	*Number of Visits*	*Mean Demandingness Rating (SD)*
Information requests[a]						
3 or fewer	48	3.24 (0.79)	20	2.94 (0.42)	68	3.16 (0.72)
4 or more	17	2.88 (0.49)	46	2.64 (0.82)	63	2.71 (0.74)
Action requests[b]						
0	37	3.28 (0.81)	30	2.81 (0.79)	67	3.08 (0.83)
1 or more	28	2.96 (0.59)	36	2.67 (0.69)	64	2.80 (0.66)
Total requests[c]						
4 or fewer	50	3.23 (0.78)	19	3.0 (0.49)	69	3.17 (0.71)
5 or more	15	2.87 (0.52)	47	2.62 (0.79)	62	2.68 (0.74)

[a] Main effect of request "count" (high vs. low) significant, $p = 0.013$. Main eect of visit length borderline significant, $p = 0.109$. Interaction of request count and visit length not significant, $p = 0.95$.
[b] Main effect of request count significant, $p = 0.025$. Main eect of visit length borderline significant, $p = 0.068$. Interaction of request count and visit length not significant, $p = 0.85$.
[c] Main effect of request count borderline significant, $p = 0.084$. Main eect of visit length significant, $p = 0.005$. Interaction of request count and visit length not significant, $p = 0.53$.

situation may or may not apply in analysis of patient–doctor discourse. For example, in the Davis Observation Code (Callahan and Bertakis 1991), the unit of analysis is a fixed (15-second) period of time. In this case, identification of the "unit" is reliable by definition. Confusion can arise when the unit of analysis is less clear cut. The Roter Interaction Analysis System (Roter, Stewart, Punam, et al. 1997) applies codes to every patient or doctor "utterance," defined as the shortest segment of speech constituting a complete thought. Raters may disagree when one utterance ends and another begins. The situation is especially acute when the coding system operates functionally in two stages, as when a segment of speech is first identified as a request and then coded as a particular type of request. Examining both unitizing and classifying dimensions of interrater agreement provides a more balanced overall assessment of reliability.

In comparing the distribution of patient requests by specialty, several similarities as well as differences emerge. The mean number of requests was similar in internal medicine and cardiology. Requests for medication refills and for information about drug therapy (medications) were common in both specialty practices. On the other hand, reflecting differences in the spectrum of care provided by the two specialties, cardiology patients asked more questions about invasive procedures, and internal medicine patients made more requests for specific components of the physical exam. These findings comport with clinical intuition (cardiologists perform more procedures and internists examine more organ systems) and are therefore unlikely to be an epiphenomenon reflecting differences in the timing of data collection.

The expanded TORP described in this article was developed through an iterative process that integrated clinical judgment with a standard empirical card-sorting technique, followed by multiple rounds of review and revision. We believe that application of this process by a multidisciplinary panel that included an internist, a family physician, a cardiologist, and a health psychologist has produced a taxonomy with substantial face and content validity. As expected, patient requests were empirically correlated with greater health worry and more physician-perceived visit demandingness. In addition, request fulfillment was associated with greater patient satisfaction. Correlations were modest, however, supporting the clinically obvious interpretation that other factors besides request fulfillment influence satisfaction. Such factors may include patient demographics and severity of illness; physician practice style apart from handling of patient requests; and the manner in which patient requests are addressed

(not just whether or not they are fulfilled). In addition, perceived request fulfillment may be a more powerful predictor of satisfaction than actual fulfillment as captured objectively on audiotape. Additional research, including qualitative inquiry directed at identifying characteristics of successful and unsuccessful clinical negotiations, is needed.

This study has several limitations affecting interpretation of the results. Because we studied 131 patients seeing only 11 physicians in a single geographic region, our estimates of request frequency cannot be generalized to other physicians or practice settings. However, our findings concerning TORP's reliability, validity, and utility are likely to be robust across practices. Although TORP can be applied to audiotape or videotape in real time, it is a complex system that requires considerable training for reliable application. This is unavoidable given the complexity of the medical encounter itself. A related limitation is that TORP cannot account for nonverbal behavior, which may be an important component of the clinical negotiation. Finally, this study did not examine patients' perceptions of their own requests or physicians' perceptions of patients' requests: it did not link requests to outcomes, and it did not look at the relational context in which requests are made. These issues are the focus of research now underway.

In summary, the refined TORP shows evidence of both unitizing and classification reliability. In addition, the system appears applicable to both generalist and specialist practices. TORP should be a useful tool for researchers seeking to understand the clinical negotiation in the context of a changing health care system.

NOTE

1. The properties of kappa are such that it is generally not very sensitive to the exact number of negative agreements under circumstances such are observed here with unitizing requests. It can be demonstrated that as the number of negative agreements becomes large relative to the number of disagreements (as is the case in the present study), kappa becomes stable. In particular, kappa approaches from below the ratio of the number of positive agreements to the average of the number of statements labeled as requests by each of the two raters individually. In other words, if n_1 is the number of requests (positive ratings) identified by rater one, n_2 is the number of requests (positive ratings) identified by rater two, and $n_{12} < \max(n_1, n_2)$ is the number of items identified as requests by both (positive agreements), kappa approaches from below the quantity: $n_{12}/[(n_1 + n_2)/2]$.

REFERENCES

Bartlett, D. F. 1999. "The New Health Care Consumer." *Journal of Health Care Finance* 25: 44–51.

Callahan, E. J., and K. D. Bertakis. 1991. "Development and validation of the Davis Observation Code." *Family Medicine* 23: 19–24.

Charon, R., M. G. Greene, and R. D. Adelman. 1994. "Multi-Dimensional Interaction Analysis: A Collaborative Approach to the Study of medical Discourse." *Social Science Medicine* 39 (7): 955–65.

Eisenthal, S., C. Koopman, and J. D. Stoeckle. 1990. The nature of patients' requests for physician help. *Academy of Medicine* 65: 401–5.

Gallagher, T. H., B. Lo, M. Chesney, and K. Christensen. 1997. "How Do Physicians Respond to Patient's Requests for Costly, Unindicated Services?" *Journal of General Internal Medicine* 12 (11): 663–8.

Garvin, B. J., C. W. Kennedy, and K. N. Cissna. 1988. "Reliability in Category Coding Systems." *Nursing Research* 37: 52–5.

Greenfield, S., S. H. Kaplan, and J. E. Ware. 1985. "Expanding Patient Involvement in Care: Effects on Patient Outcomes." *Annals of Internal Medicine* 102: 520–8.

Kravitz, R. L., R. A. Bell, and C. E. Franz. 1999. "A Taxonomy of Requests by Patients (TORP): A New System for Understanding the Clinical Negotiation in Office Practice." *Journal of Family Practice* 48: 872–8.

Kravitz, R. L., S. Greenfield, and W. Rogers, W. G. Manning, Jr., M. Kubkoff, E. C. Nelson, A. R. Tarlov, W. E. Ware, Jr. 1992. "Differences in the Mix of Patients Among Medical Specialties and Systems of Care. Results from the Medical Outcomes Study." *Journal of the American Medical Association* 267: 1617–23.

Kravitz, R. L., E. J. Callahan, R. Azari, D. Antonius, and C. E. Lewis. 1997. Assessing patients' expectations in ambulatory medical practice. Does the measurement approach make a difference? *Journal of General Internal Medicine* 12 (1): 67–72.

Larsen, J. H., O. Risor, and S. Putnam. 1997. "P-R-A-C-T-I-C-A-L: A Step-by-Step Model for Conducting the Consultation in General Practice." *Family Practice* 14: 295–301.

Lazare, A., S. Eisenthal, and L. Wasserman. 1975. "The Customer Approach to Patienthood: Attending to Patient Requests in a Walk-in Clinic." *Archives of General Psychiatry* 32: 553–8.

Like, R., and S. J. Zyzanski. 1986. "Patient Requests in Family Practice: A Focal Point for Clinical Negotiation." *Family Practice* 3: 216–28.

MacKenzie, E. J., D. M. Steinwachs, L. R. Bone, D. J. Floccare, and A. I. Ramzy. 1992. "Inter-Rater Reliability of Preventable Death Judgments. The Preventable Death Study Group." *Journal of Trauma* 33: 292–302.

Roter, D., and J. A. Hall. 1992. *Doctors Talking with Patients/Patients Talking with Doctors: Improving Communication in Medical Visits.* Westport, CT: Auburn House.

Roter, D. L., M. Stewart, S. Punam, M. Lipkin, W. Stiles, and T. Inui. 1997. "Communication Patterns of Primary Care Physicians." *Journal of the American Medical Association* 277: 350–5.

StataCorp. 1999. *Stata Statistical Software* (release 6.0). College Station, TX: Stata Corporation.

Ukoumunne, O. C., M. C. Gulliford, S. Chinn, J. A. C. Sterne, and P. G. J. Burney. 1999. "Methods for Evaluating Area-Wide and Organisation-Based Interventions in Health and Health Care: A Systematic Review." *Health Technology Assessment* 3 (5).

Ware, J. E. Jr., and R. D. Hays. 1988. "Methods for Measuring Patient Satisfaction with Specific Medical Encounters." *Medical Care* 26: 393–402.

Wasserman, R. C., and T. S. Inui. 1983. "Systematic Analysis of Clinician-Patient Interactions: A Critique of Recent Approaches with Suggestions for Future Research." *Medical Care* 21: 279–92.

Weijer, C., P. A. Singer, B. M. Dickens, and S. Workman. 1998. "Dealing with Demands for Inappropriate Treatment." *Canadian Medical Journal Association* 159: 817–21.