The Doctor-Patient Relationship: Can it Be Measured? A Review of the Measurement of the Interpersonal Aspects of Care — Prof Ronald J Henbest

Summary
The importance of the doctor-patient relationship has been increasingly recognized, especially during the past three decades. With this recognition has followed the need for adequate measurement instruments with which to study it.

This paper reviews 8 methods that have been designed to measure the doctor-patient interaction in family practice. In addition, to provide appropriate background and perspective, two other groups of instruments are described; namely, those developed by the client-centred therapists and those commonly referred to as systematic interaction analysis methods. Both sets of instruments are older and have been used extensively. However, they were not designed specifically for the context of the doctor-patient interaction and are shown to be inadequate for its measurement.

Thus, there is a need to develop new methods to describe and analyze the doctor-patient interaction in family practice. Some of the newer methods have demonstrated validity and reliability and one has even shown an association between the doctor-patient interaction and patient outcomes. The work continues.

Introduction
The realization that the doctor-patient relationship is of tremendous importance is not a new one. In 1927 Sir Francis Peabody was so bold as to declare that: "The significance of the intimate personal relationship between physician and patient cannot be too strongly emphasized, for in an extraordinarily large number of cases both diagnosis and treatment are directly dependent on it, and the failure of the young physician to establish this relationship accounts for much of his ineffectiveness in the care of patients".

Forty years later the following statement appeared in an editorial in the Lancet: "Care of the doctor-patient relation has for too long been left to chance; because of its importance to general practice it must now be examined, defined, and taught, for only then can it be practiced effectively." Indeed, during the past three decades much examination, definition and teaching of the doctor-patient relationship has occurred; and the need to demonstrate a relationship between process and outcomes has been emphasized. But to do that, we need to be able to measure the doctor-patient interaction.

The primary purpose of this paper is to review the methods that have been developed specifically to measure the doctor-patient interaction in family practice and the reader may wish to turn directly to that section. However, in order to provide appropriate background and perspective, I shall first briefly describe two older and widely used groups of instruments; namely, those developed by the client-centred therapist and those produced by the systematic interaction analysts.

---

**Prof Ronald J Henbest**
MD, CCFP, M CI Sc (Fam Med)
The Dept of Family Medicine; Medunsa PO Box 222 Medunsa, 0204

Curriculum vitae
Ronald J Henbest was born in Edmonton, Alberta (Canada) where he qualified in 1974 with a BSc in Maths and Psychology and in 1978 with an MD from the University of Alberta. He then completed two years' postgraduate study (residency) in Family Medicine at the University of Western Ontario (Canada) and obtained his CCFP from the College of Family Physicians of Canada. Ron joined the Department of Family Medicine at Medunsa in 1980. He has a particular interest in the doctor-patient interaction and its importance for healing. He returned to the University of Western Ontario in 1984 to take their Master of Clinical Science Degree in Family Medicine (MCIsCe), which emphasizes patient care, teaching and learning, and research. His thesis on Patient Centred Care involved the development of a method for measuring patient-centredness and testing it against patient outcomes. In 1989 Ron returned to his home city, Edmonton, for a period of 21 months where he was engaged as an associate professor in the Department of Family Medicine at the University of Alberta. During this time, he also completed further training in systemic family therapy. In October 1990, Ron returned, with his wife Judy and four year old son Benji, this time as associate professor and deputy head of the Department of Family Medicine at Medunsa.
The use of these scales, particularly those of the core dimensions has been extensive. Most typically, the scales have been applied to brief samples of three minutes duration excerpted from audio-tape recordings of psychotherapy. Reliabilities for the three core dimensions (expressed as Pearson Correlations) have ranged from 0.42 to 0.79 for the empathy scale, from 0.23 to 0.84 for the unconditional positive regard scale, and from 0.20 to 0.62 for the genuiness scale as reported by Truax and Mitchell in a review of more than thirty-five studies.14

At present, the measurement scales are being used in a modified form in human development programs.19,20

These scales have not to my knowledge been applied to any great extent to the doctor-patient interaction in family practice, although Lehman used the empathy, warmth, and genuiness scales as outcome measures in a study called, "Intake History and Physical Examination and Its Association with the Doctor-Patient Relationship".21

The use of these scales, particularly those of the core dimensions has been extensive. Most typically, the scales have been applied to brief samples of three minutes duration excerpted from audio-tape recordings of psychotherapy. Reliabilities for the three core dimensions (expressed as Pearson Correlations) have ranged from 0.42 to 0.79 for the empathy scale, from 0.23 to 0.84 for the unconditional positive regard scale, and from 0.20 to 0.62 for the genuiness scale as reported by Truax and Mitchell in a review of more than thirty-five studies.14

At present, the measurement scales are being used in a modified form in human development programs.19,20

These scales have not to my knowledge been applied to any great extent to the doctor-patient interaction in family practice, although Lehman used the empathy, warmth, and genuiness scales as outcome measures in a study called, "Intake History and Physical Examination and Its Association with the Doctor-Patient Relationship".21

The use of these scales, particularly those of the core dimensions has been extensive. Most typically, the scales have been applied to brief samples of three minutes duration excerpted from audio-tape recordings of psychotherapy. Reliabilities for the three core dimensions (expressed as Pearson Correlations) have ranged from 0.42 to 0.79 for the empathy scale, from 0.23 to 0.84 for the unconditional positive regard scale, and from 0.20 to 0.62 for the genuiness scale as reported by Truax and Mitchell in a review of more than thirty-five studies.14

At present, the measurement scales are being used in a modified form in human development programs.19,20

These scales have not to my knowledge been applied to any great extent to the doctor-patient interaction in family practice, although Lehman used the empathy, warmth, and genuiness scales as outcome measures in a study called, "Intake History and Physical Examination and Its Association with the Doctor-Patient Relationship".21

The use of these scales, particularly those of the core dimensions has been extensive. Most typically, the scales have been applied to brief samples of three minutes duration excerpted from audio-tape recordings of psychotherapy. Reliabilities for the three core dimensions (expressed as Pearson Correlations) have ranged from 0.42 to 0.79 for the empathy scale, from 0.23 to 0.84 for the unconditional positive regard scale, and from 0.20 to 0.62 for the genuiness scale as reported by Truax and Mitchell in a review of more than thirty-five studies.14

At present, the measurement scales are being used in a modified form in human development programs.19,20

These scales have not to my knowledge been applied to any great extent to the doctor-patient interaction in family practice, although Lehman used the empathy, warmth, and genuiness scales as outcome measures in a study called, "Intake History and Physical Examination and Its Association with the Doctor-Patient Relationship".21

The use of these scales, particularly those of the core dimensions has been extensive. Most typically, the scales have been applied to brief samples of three minutes duration excerpted from audio-tape recordings of psychotherapy. Reliabilities for the three core dimensions (expressed as Pearson Correlations) have ranged from 0.42 to 0.79 for the empathy scale, from 0.23 to 0.84 for the unconditional positive regard scale, and from 0.20 to 0.62 for the genuiness scale as reported by Truax and Mitchell in a review of more than thirty-five studies.14

At present, the measurement scales are being used in a modified form in human development programs.19,20

These scales have not to my knowledge been applied to any great extent to the doctor-patient interaction in family practice, although Lehman used the empathy, warmth, and genuiness scales as outcome measures in a study called, "Intake History and Physical Examination and Its Association with the Doctor-Patient Relationship".21

The use of these scales, particularly those of the core dimensions has been extensive. Most typically, the scales have been applied to brief samples of three minutes duration excerpted from audio-tape recordings of psychotherapy. Reliabilities for the three core dimensions (expressed as Pearson Correlations) have ranged from 0.42 to 0.79 for the empathy scale, from 0.23 to 0.84 for the unconditional positive regard scale, and from 0.20 to 0.62 for the genuiness scale as reported by Truax and Mitchell in a review of more than thirty-five studies.14

At present, the measurement scales are being used in a modified form in human development programs.19,20

These scales have not to my knowledge been applied to any great extent to the doctor-patient interaction in family practice, although Lehman used the empathy, warmth, and genuiness scales as outcome measures in a study called, "Intake History and Physical Examination and Its Association with the Doctor-Patient Relationship".21

The use of these scales, particularly those of the core dimensions has been extensive. Most typically, the scales have been applied to brief samples of three minutes duration excerpted from audio-tape recordings of psychotherapy. Reliabilities for the three core dimensions (expressed as Pearson Correlations) have ranged from 0.42 to 0.79 for the empathy scale, from 0.23 to 0.84 for the unconditional positive regard scale, and from 0.20 to 0.62 for the genuiness scale as reported by Truax and Mitchell in a review of more than thirty-five studies.14

At present, the measurement scales are being used in a modified form in human development programs.19,20

These scales have not to my knowledge been applied to any great extent to the doctor-patient interaction in family practice, although Lehman used the empathy, warmth, and genuiness scales as outcome measures in a study called, "Intake History and Physical Examination and Its Association with the Doctor-Patient Relationship".21

The use of these scales, particularly those of the core dimensions has been extensive. Most typically, the scales have been applied to brief samples of three minutes duration excerpted from audio-tape recordings of psychotherapy. Reliabilities for the three core dimensions (expressed as Pearson Correlations) have ranged from 0.42 to 0.79 for the empathy scale, from 0.23 to 0.84 for the unconditional positive regard scale, and from 0.20 to 0.62 for the genuiness scale as reported by Truax and Mitchell in a review of more than thirty-five studies.14

At present, the measurement scales are being used in a modified form in human development programs.19,20

These scales have not to my knowledge been applied to any great extent to the doctor-patient interaction in family practice, although Lehman used the empathy, warmth, and genuiness scales as outcome measures in a study called, "Intake History and Physical Examination and Its Association with the Doctor-Patient Relationship".21

The use of these scales, particularly those of the core dimensions has been extensive. Most typically, the scales have been applied to brief samples of three minutes duration excerpted from audio-tape recordings of psychotherapy. Reliabilities for the three core dimensions (expressed as Pearson Correlations) have ranged from 0.42 to 0.79 for the empathy scale, from 0.23 to 0.84 for the unconditional positive regard scale, and from 0.20 to 0.62 for the genuiness scale as reported by Truax and Mitchell in a review of more than thirty-five studies.14

At present, the measurement scales are being used in a modified form in human development programs.19,20

These scales have not to my knowledge been applied to any great extent to the doctor-patient interaction in family practice, although Lehman used the empathy, warmth, and genuiness scales as outcome measures in a study called, "Intake History and Physical Examination and Its Association with the Doctor-Patient Relationship".21

The use of these scales, particularly those of the core dimensions has been extensive. Most typically, the scales have been applied to brief samples of three minutes duration excerpted from audio-tape recordings of psychotherapy. Reliabilities for the three core dimensions (expressed as Pearson Correlations) have ranged from 0.42 to 0.79 for the empathy scale, from 0.23 to 0.84 for the unconditional positive regard scale, and from 0.20 to 0.62 for the genuiness scale as reported by Truax and Mitchell in a review of more than thirty-five studies.14

At present, the measurement scales are being used in a modified form in human development programs.19,20

These scales have not to my knowledge been applied to any great extent to the doctor-patient interaction in family practice, although Lehman used the empathy, warmth, and genuiness scales as outcome measures in a study called, "Intake History and Physical Examination and Its Association with the Doctor-Patient Relationship".21
interaction (gives suggestion, opinion, orientation, asks for suggestion, opinion, orientation). Reliability of this method has been found to be high with 91.6% inter-observer agreement for unitizing and 82.4% agreement for categorizing. Bales' system deals best with the relationship aspect of the encounter in that it is sensitive to the feelings of the doctor and patient toward each other. However, its classification of information transfer is inadequate for the doctor-patient interaction and the system as a whole does not take context into account.

Roter's method for interactional analysis is a modification of Bales' System that was originally devised to study patient question-asking behaviour. Eight categories are used to document information transfer; affect is measured by a global rating scale. As with Bales' system, the immediate context of the interaction is not measured and sequence is not taken into account.

Stiles' Verbal Response Mode System was designed specifically for the analysis of dyadic exchanges. His taxonomy of verbal responses is meant to define a mutually exclusive and exhaustive set of eight basic modes. Each mode is intended to describe the particular function that the verbal act was meant to achieve. The unit of analysis is the utterance, which Stiles defines as "the grammatical equivalent of one psychologic unit of experience." This system best describes the ways in which information is transferred in the encounter; affect is omitted as is any analysis of content or context of the interaction.

Katz's system of Resource Exchange Analysis defines resources as important goods or services (such as information, greeting, diagnosis), and categorizes the modes of exchange as being either initiating or responding. The unit of analysis is the interact, which consists of the initiation of a response by one communicator and the response to it by the other. This system emphasizes the content and the social context of the doctor-patient interaction and has the additional advantage of a conceptually appealing unit of analysis. The interact potentially allows for analysis of interaction sequence. Problems with the resource exchange system include difficulty in the designation of interacts and the fact that some interaction sequences do not fit easily into the resource exchange model of initiating and response modes. Two recent reviews of systematic interaction analysis have emphasized the need for the development of new ways of analyzing consultations that are specific for the context of the interaction. None of the representative systems of analysis reviewed (Bales', Roter's, Stiles' and Katz's) are ideally suited to the doctor-patient encounter. None of them (for example, adequately identifies a patient's effort to raise an issue which may be critical for determining the patient's real reason for attendance. This would be a particularly serious deficiency for an instrument being used to assess the interpersonal process in family practice.

The Doctor-Patient Interaction: Its Measurement in Family Practice

I shall now examine 8 specific attempts to describe and analyze the doctor-patient interaction in family practice.

(1) Balint et al, noting the very complex difficulties inherent in attempting to rate the doctor-patient relationship, limited themselves to, "a very crude assessment based on the general practitioner's own judgment." For this purpose they first used a five point scale: hot war, uneasy truce, peace, friendly negotiations, and mutual trust. However, even this proved to be too complicated, and they reduced their scales to three points: negative, peace, and positive.

(2) As part of a larger study of doctor patient communication, Gozzi, Morris and Korsch devised a method of analysis to learn how two people in a medical setting help and hinder one another in expressing themselves. Three sets of categories were developed. The facilitation categories designated statements which one person made that were in accord with the preceding comments of the other person. Three sets of categories were developed. The facilitation categories designated statements which one person made that were in accord with the preceding comments of the other person. The blocking categories included statements that were not in accord with the preceding ones or statements that fit one of ten specifically
defined doctor-block categories. Examples of doctor-block categories were the doctor interrupting the patient or ignoring the patient’s comments. The third category was for interactions which could not be coded as facilitations or blocks. The rating for each doctor-patient interaction was calculated from the numbers of doctor blocks and facilitations to yield a percentage of blocks per visit. This method of analysis has been applied to a study of 82 doctor-patient interactions. 

Critc criterion validity for the method was supported in that those interactions which had scored high in doctor blocks, showed proportionately more negative affective statements by the doctor as determined by Bales' system of analysis. No indication of reliability testing was reported.

(3) Byrne and Long developed a method for a detailed analysis of the doctor's behaviour in the consultation. Their method involves the use of checklists, comprised of specific behaviours, for each of two major categories of doctor behaviour: doctor-centred behaviour and patient-centred behaviour. The doctor-centred category includes such behaviours as asking closed questions, directing, and giving information. Patient-centred behaviours include the use of broad questions, reflecting, and accepting patient ideas or feelings. A third category, labelled negative behaviour, is used to classify behaviour that rejected or denied the patient in some way.

In addition, Byrne and Long developed a scoring system that identifies a doctor’s basic consultation style. The manner in which this method of analyzing consultations was developed (involving the detailed examination of the recordings of over 2500 doctor-patient interviews) grants it considerable validity. As far as I am aware, there are no reports of others using this method in their research and no reports of reliability testing. One drawback to the scoring method is that it requires a transcript of the consultation because of the detail required by each of the two distinct checklists that have to be completed.

A consultation represents just one interaction of the many which make up the doctor-patient relationship.

(4) Bain through the detailed study of 480 audio-tape recordings of his own consultations, developed a method of analysis using ten categories (five each for physician and patient) to classify the verbal content of the doctor-patient interaction. He later increased the number of categories to six each. The categories for classifying the doctor’s verbal behaviour are: social exchange, facilitation, asking questions, medical problem resolutions (answers regarding medical problems or treatment), response to social problems, and instruction. The categories for patients are: presentation of symptoms, answering questions, response to instruction, problem-related expressions, questions, and social exchange. This method does not allow sequence to be analyzed, nor do the categories selected by Bain seem to allow for an assessment of the extent to which a consultation is patient-centred other than in the very general sense of who does the most talking.

(5) An interview rating scale for general practice has been designed and tested by Verby et al. The scale consists of 17 items. Eleven of these items could be classified as statements of interview technique, three of the items describe nonverbal aspects of communication, and the remaining three relate to verbal content. Each item is scored on a four point scale. The items for this rating scale were carefully chosen and the scale would appear to be a valid way of assessing patient-centred care. However, this method of rating a consultation does not allow for the examination of any particular interaction sequence between doctor and patient, and it requires either direct observation or the use of video recordings to assess the nonverbal items.

(6) Pendleton et al has devised a consultation rating scale based on 7 tasks for the consultation. These tasks are:

(1) to define the reasons for the patients’ attendance,
(2) to consider other problems,
(3) to choose (with the patient)
The Doctor-Patient Relationship

An appropriate action for each problem,

(4) to achieve a shared understanding of the problems with the patient,

(5) to involve the patient in the management of the problems,

(6) to use time and resources appropriately, and

(7) to establish or maintain a relationship with the patient which helps to achieve the other tasks.

Each part of each task is rated separately such that there are 14 items on the scale, each consisting of two opposing statements linked by a line. The rater places a mark on the line to show how much he agrees with the opposing statements. For example, the first item consists of the statements: nature and history of problems inadequately defined, nature and history of problems adequately defined. The 14 items provide for a comprehensive assessment of the consultation taking into account both the interpersonal and the technical skills of the physician. This scale is specifically intended to provide feedback to trainees. The theoretical base for the instrument is well articulated. No studies have been reported with it to date.

Recently, the patient-centred clinical model developed at the University of Western Ontario has been defined in operational terms and a method devised specifically for scoring patient-centredness. This method would qualify as a system of interaction analysis as defined by Inui et al. The observational strategy employed may be that of direct observation, audio-tape, or audio-video recording. Its process of particular interest is that of the doctor's verbal response to what the patient says. Its taxonomy for categorizing encounter events involves firstly, classifying the patient's communications as expectations, feelings, fears and prompts, and secondly classifying the doctor's response as to whether he acknowledges (or not) and cuts off (or not) the patient's communication. In addition, the facilitating behaviours of the physician are listed. All the terms used for categorizing events are operationally defined. The scoring system involves assigning a score from 1 to 4 to each of the five categories of physician behaviour; namely, his responses to each of the four categories of patient communications (expectations, feelings, fears and prompts) and his facilitating behaviours which are listed separately. This method would appear to be the most specific method thus far developed for measuring the concept of patient-centredness in that it focuses directly on the patient's agenda, as defined in terms of patient expectations, feelings and fears. It also provides for the analysis of sequence to some extent, in that the patient's communication is recorded verbatim and then the doctor's response to that communication is classified.

However, this method is still in its developing stages and needs further work. Firstly, more categories for classifying patient communications would be helpful (for example, there is no provision for classifying a patient's thoughts about his illness). A second related problem is that the expectation category is defined very broadly (it includes everything that is not a feeling or a fear). Thirdly, a broader range of categories of doctor responses representing varying degrees of patient-centredness might allow the instrument greater sensitivity in differentiating between physicians with regard to this factor. A fourth problem is that different raters recording different patient statements so that it is different aspects of the same interview that are being scored. A fifth area that needs attention is the mechanics of the scoring system. As it stands, each category of patient behaviour is given equal weight so that different numbers of patient offers in different categories results in unequal weighting of those offers. Because it is very difficult to weigh individual patient cues according to their potential significance, it would seem better, at least for the present, to assign equal weight to the doctors response to each of these cues.

Non-verbal behaviour is an important part of the doctor-patient communication.

(8) A modification of the method piloted by the University of...
Western Ontario, Henbest et al. has attempted to solve some of these problems and in an initial trial, has demonstrated not only validity and reliability but also an association with patient outcomes.

This method assessed the doctor-patient interaction specifically in terms of its patient-centredness with patient-centredness defined as care in which the doctor responds to the patient in such a way as to facilitate the patient's expression of all his or her reasons for coming to the doctor, including symptoms, expectations, thoughts, and feelings. A score sheet is used to record the offers (any potentially significant response) made by the patient and to score the doctor's response(s) to each offer on a 4 point scale as follows: (0) if the doctor ignored the offer altogether, (1) if closed responses were used, (2) if open-ended responses were given, and (3) if the doctor specifically facilitated the expression of the patient's expectations, thoughts, or feelings. In addition to demonstrating validity and reliability, this method has also shown sensitivity in that the patient-centred scores varied significantly among the practitioners studied and also varied significantly for the responses to different categories of patient offers; symptoms received the most patient-centred responses followed by thoughts and expectations, with feelings receiving the least patient-centred responses. The method was also found to be practical in that it was cheap, could be used in a variety of situations, and showed potential for being very time effective as the score for the first two minutes of the consultation was highly correlated with the patient-centred score for the entire consultation ($r_s = 0.806, p = 0.001$). Perhaps most important of all, this method has been able to demonstrate that the doctor's responses to a patient do make a difference. Patient-centredness was found to be associated with the doctor having ascertained the patients reasons for coming and with resolution of the patient's concerns.

Discussion

The measurement of the doctor-patient relationship will be discussed in terms of three major properties: validity, reliability, and practicality.

Validity

Four main approaches to the study of validity are commonly distinguished: face validity, content validity, concurrent or criterion validity, and construct validity.

It would seem that for a method for analyzing the consultation in family practice to have face validity, that three things would have to be taken into account. First, the categories used for classifying and scoring behaviours should be appropriate to the doctor-patient interaction. Second, the method should allow for interaction sequence to be identified and evaluated. In order to assess the appropriateness of what the practitioner says or does it is important to know what has immediately gone on before. For example, a statement that sounds facilitating when taken in isolation, may be entirely inappropriate when considered in context. Third, the context of the interaction itself; that is, as part of an ongoing doctor-patient relationship, needs to be taken into account.

The measurement instruments used by the client-centred therapists are inadequate for assessing the family practice consultation primarily because they do not cover the full range of family doctor behaviours or responsibilities. The instruments developed by the interaction analysts also lack face and content validity in that the categories used are not specific for the doctor-patient interaction and on the whole, they do not allow for the assessment of the doctor's specific response to a patient's specific offer or complaint.

The newer methods, developed specifically to assess the doctor-patient interaction, appear to have more appropriate categories for classifying behaviours as well as broader ranges of items assessed. The categories used by Byrne and Long, Bain, Verby et al, Pendleton et al, and the University of Western Ontario (UWO), all seem well suited to the family practice consultation and cover most of its content. However, only the methods described by Gozzi et al. and Henbest et al. allow for interaction sequence to be taken into account. The UWO method has demonstrated criterion validity using appropriate categories of Bales' System for comparison.

The method developed by Henbest et al. in initial trials, has also demonstrated criterion validity, both with the UWO method and with the Empathy Scale used by the client-centred therapists. In addition, it has demonstrated construct validity.
in that patients who had experienced interviews that were assessed as being more patient-centred were significantly more likely to report that their reasons for coming had been ascertained and that they had felt really understood by the doctor, than patients who had experienced interviews assessed as less patient-centred.

An important, but difficult reality to take into account, is the fact that most often in family practice, a consultation represents just one interaction of many that make up the doctor-patient relationship. None of the methods described in this paper provide a measure of the overall relationship.

One further validity issue requires mention, the measurement of nonverbal as well as verbal communications. Nonverbal behaviour is recognized to be an important part of communication, but much more work is necessary in order to develop a method to assess it in the context of the doctor-patient interaction. Only the method described by Verby et al takes nonverbal behaviours into account and only to a small extent.

Reliability

Reliability or reproducibility is a measure of agreement between repeated measurements on the same subject. Two aspects would seem important here:

1. the recording of the consultation, and
2. the scoring of the consultation.

Methods that require recording of the consultation in some way include those described by Byrne and Long, Bain, UWO, and Henbest et al. Bain provides some evidence for the reliability of his method by showing that two independent raters had put the same number of units of expression in each of five categories, but it is not clear that the same units of expression were placed in the same categories. Henbest et al found that two raters independently recorded over 80% of the patient offers identified by the other rater. The main difference between the raters was the extent to which they grouped the patients offers, especially symptoms, together. Reliability of the scoring of consultations has been reported for the method described by Verby et al (Pearson Correlation Coefficients 0.87 and 0.80 for inter- and intra-rater reliabilities respectively) UWO (Pearson r = 0.687; 0.835; and 0.803 for inter-rater reliabilities), and for Henbest et al (Spearman Rank Correlation Coefficients rs = 0.91; p = 0.001; and rs = 0.88; p = 0.002 for inter- and intra rater reliabilities respectively).

Practicality

The main practical considerations are the resources and time required for measurement. The measurement of nonverbal behaviour, for example Verby et al, requires either direct observation or video equipment. The completion of detailed checklists such as those compiled by Byrne and Long requires transcripts. The other methods presented can be scored either from direct observation or from audio-tapes. The time required for scoring either matches the length of the consultation if direct observation is used or may be longer if audi-tapes are used (which allows for replay for careful scrutiny of the interaction). Of note, the method developed by Henbest et al, in an initial trial, showed a high positive correlation between the patient-centred score determined from the first two minutes of the audiotape of the consultation and the score determined from the entire tape of the consultation. The use of a two-minute score would prove very useful especially in large scale studies.

Conclusions

Encouraging progress has been made towards developing practical, valid and reliable measurement instruments. At present, no methods are available for measuring the total doctor-patient relationship, rather than single consultations.

In 1964, Pellegrino, in an article titled, "Patient Care - Mystical Research or Researchable Mysterie?", stated that: “Investigators seem to have settled for what is measurable instead of ..."
measuring what they would really like to know."

The study of the doctor-patient relationship is an important task, but not an easy one. Let us not settle for simply measuring the measurable. Rather, let us continue to strive towards developing the means for measuring what we would really like to know, namely, what makes for an effective doctor-patient interaction, one that is satisfying for both doctor and patient and that leads to improved patient outcomes?

Acknowledgements
I would like to thank Prof Mike Brennan, Prof Wayne Weston, Dr Moira Stovart, Prof Martin Bass, and Prof Sam Fehsen for many helpful discussions concerning the doctor-patient interaction.

References
1. Peabody FW. The Care of the Patient. JAMA 1927; 88: 877.
32. Sillies WB, Putnam SM, Wolf MH, James SA. Verbal Response Mode Profiles of
... The Doctor-Patient Relationship


