

Description versus evaluation of medical interviews

WILLIAM B. STILES

Department of Psychology, Miami University, Miami, USA

In many areas of medicine, an intervention can be evaluated using the same measures and terms as are used to describe it. What is good - that is, what is effective in producing medical benefit - can be characterized as specifiable amounts of particular actions or substances delivered under specifiable conditions. Research that develops a more precise description of the intervention leads directly to a more precise description of what should be done to maximize benefit.

The situation is different in the case of interviewing. As illustrated in the other editorials in this series (Rimondini & Del Piccolo, 2002; Putnam, 2002), precise descriptions of the interview process often fail to yield comparably precise specifications of how interviews should be conducted. What is happening here? In this editorial, I suggest that the complicating factor is the intricate responsiveness of human interaction, in which the participants continually adjust their actions to achieve desirable outcomes. Although it is possible to describe interviews with considerable precision and to evaluate interviews with acceptable reliability, different measures are used for description and evaluation. I will try to explain why.

WHAT IS RESPONSIVENESS?

Responsiveness describes behavior being affected by emerging context, including shifting perceptions of others' characteristics and behavior (Stiles *et al.*, 1998). Human interaction is responsive on time scales that range from milliseconds to months. In conversations,

people answer each other's questions, stay on related topics, take turns speaking using an elaborate system of signals (e.g., Frankel, 1984; Goodwin, 1981; Grice, 1975; Sacks *et al.*, 1974). In medical interviews, responsiveness may include treatment selection and planning based on presenting problems and characteristics as well as the timing and phrasing of interview components, based on judgments about patients' level of understanding and physical or emotional state. Patients present with different requirements, and the requirements change from moment to moment as their interviews progress. Responsiveness implies that the participants' behavior changes as the requirements change. Insofar as physician and patient respond to each other, there is a dynamic relationship between interview variables, involving bi-directional causation and feedback loops. The content and process emerge as the interview proceeds, rather than being planned completely in advance. Thus, no two patients receive identical interviews. The term responsiveness is technically neutral. In some contexts, the intent of responsive behavior could be hostile or destructive. Our primary interest, however, is in appropriate responsiveness, doing what is required to produce some positive, desired outcome, such as patient satisfaction, adherence, reduction in symptoms, or improvement in life functioning. In addition to their normal social competence and human caring, physicians' training and supervision promote appropriate responsiveness. The physician's job is to assess the patient and the circumstances and to choose a treatment that is appropriate for the patient's problems, follow a strategy that is appropriate for the patient's capacity, and take actions appropriate for the patient's current state.

Appropriate *responsiveness* thus demands sensitivity both to emerging patient requirements and to the goals and tasks of the interview. A physician must, to some degree, recognize the patient's needs and problems and

Address for correspondence: Professor W.B. Stiles, Department of Psychology, Miami University, Oxford, OH 45056 (USA).

Fax: +1-513-529.2420

E-mail: stileswb@muohio.edu

the patient's resources (e.g., ability to make use of) as they emerge and must respond in ways that take them into account. Patient requirements and appropriate responsiveness must be judged in relation to the setting, presenting problem, and the treatment's goals and standards. Viewed in this way, a patient's requirements may differ from his or her momentary wishes, and sometimes good treatment may involve refusing to comply with a patient's requests.

Responsiveness in medical interviews draws heavily on general human social intelligence. Not everybody does it equally well, of course, and physicians have sometimes been criticized for failure to be optimally responsive to their patients' needs. Moreover, responsiveness is undoubtedly degraded by boredom, by exhaustion, and perhaps by large differences in social status between physician and patient. Physicians should be careful on all of these scores. Nevertheless, it is important to keep in mind a comparison with complete unresponsiveness: Imagine a physician entering an interview and repeating exactly the words, behaviors, timing, and so forth of the preceding interview, regardless of what the patient does. Relative to such a truly unresponsive comparison point, even seemingly insensitive physicians must be considered as highly responsive.

WHY IS RESPONSIVENESS A PROBLEM?

My attention was drawn to the responsiveness problem by the difficulties encountered during a long search for correlations between verbal processes, such as asking questions, and the outcomes of medical and psychotherapeutic interviews (Stiles, 1989; in press). Originally, like many other researchers, my collaborators and I had reasoned more or less as follows: If an interview process component, such as physicians' question-asking or information-giving, is important, then patients who received more of that component should tend to have better outcomes than those who received less, so process and outcome should be positively correlated across patients (McDaniel *et al.*, 1981; Stiles *et al.*, 1979). Although some interesting patterns emerged, they were weaker and less consistent than we expected (e.g., Putnam *et al.*, 1985; Stiles & Shapiro, 1994). Other researchers using their own measures were similarly disappointed. At first, we thought that we might have been wrong about which components were important or that we had failed to measure the components accurately. However, I now think the main

reason for our difficulties lay in our use of linear-model statistics, such as correlation and regression, to describe responsive, and hence non-linear, processes of human interaction (Stiles, 1989; Stiles *et al.*, 1998).

The problem can be illustrated using a constructed example: Consider the effect of physician question-asking on making an accurate diagnosis for patients presenting with chest pain. Well-informed, forthcoming patients tend to provide good information in response to a few questions, whereas poorly informed, recalcitrant, or verbally unskilled patients tend to provide poorer information despite many questions. Suppose that physicians are appropriately responsive to patient differences, asking more or fewer questions depending on how forthcoming patients are. Then it is likely that the recalcitrant patients will be asked more questions but may still receive less accurate diagnoses, whereas forthcoming patients will be diagnosed more accurately despite having been asked fewer questions. That is, questioning and diagnostic accuracy may be *negatively* correlated despite the *positive* causal contribution of questioning to diagnosis. Reviewers seeing negative correlations across several studies would be likely to conclude, erroneously, that questioning is useless or counterproductive.

One can construct parallel examples for virtually any process component or outcome index. Furthermore, experimental interventions can similarly be vitiated by responsiveness, as participants compensate for the intervention by responding appropriately using alternative behaviors (Stiles, 1989; Stiles *et al.*, 1998; see also reviews of studies by Miller *et al.*, 1993, and Williams & Deci, 2001, in the accompanying editorial by Putnam (2002), for examples of interventions that appeared to make no difference). The problem is using linear models to assess process-outcome relationships that are responsive and hence obviously not linear. Information about probable outcome – how the patient is responding – continually feeds back to influence the interview process (e.g., if a question is adequately answered, the physician does not ask it again). That is the whole point of skillful interviewing. We cannot expect to find linear relationships such as, “more questioning promotes accurate diagnosis,” because the independent and dependent variables interact with each other.

The responsiveness logic thus implies that crucially important process components (such as questioning) may nevertheless have null correlations with outcomes. Of course process-outcome correlations are not always null (Putnam, 2002, reviews some that were not).

However, because they *may* be null (or negative) even when a component is crucially important, they are always untrustworthy indexes of a component's value.

DESCRIPTIONS CAN BE USEFUL

The responsiveness problem is *not* a measurement problem. The problem would remain even if we had perfect measures of key process components and outcomes. In fact, researchers, including the authors of this series of editorials, have developed a variety of reliable and precise measures for describing medical interviews (e.g., Del Piccolo *et al.*, 1998; 2000; 2002; Putnam & Stiles, 1993; Stiles, 1992; Stiles & Putnam, 1992). To illustrate, some of these instruments rate physician and patient speech into categories of speech acts, such as open or closed questions, information-giving, or instructions, or into categories dealing with specific content, such as particular body systems, health history, emotions, or social background. Such classifications and descriptions are important scientifically, to identify the elements, structure, and sequence of the process of interviewing.

Descriptive studies can inform clinical practice. For example, factor analysis of descriptive codes has shown that, across a wide variety of types of patients, diagnoses, and settings, the medical history-taking segment (typically the first part of the interview) consists mainly of two types of verbal exchanges between physician and patient: (a) *closed question exchanges*, in which physicians ask questions and patients give brief (mainly yes/no) answers, and (b) *exposition exchanges*, in which patients give tell their concerns and give background information while physicians encourage them with brief acknowledgments such as "mm-hm" (Stiles, 1996; Stiles & Putnam, 1992). Teachers of medical interviewing have traditionally emphasized using open-ended questions to identify the patient's concerns followed by closed-ended questions to collect clinical information. The descriptive coding research has usefully shown that open questions play a minor role in exposition exchanges. Indeed, peppering patients with many open questions may be counterproductive. Instead, exposition is facilitated by allowing space and active listening (Putnam *et al.*, 1988). Recasting the traditional emphasis as balancing exposition exchanges with closed question exchanges shifts the focus from the clinician's behavior (asking particular types of questions) to the joint responsive behavior of patient and physician and to the kind of information being exchanged (Putnam &

Stiles, 1993). Interestingly, an additional, type of exchange was observed in general practice interviews with patients who presented psychological problems (Cape & Stiles, 1998). In contrast to the usual unitary exposition exchange that incorporated both objective and subjective information, these patients tended to engage in an additional *emotional exposition* type of exchange, which involved mainly revealing subjective information, that is, the patient's thoughts and feelings. With its two distinct types of exposition exchange, the exchange structure of these interviews somewhat resembled that of psychotherapy interviews (Stiles & Shapiro, 1995), presumably reflecting a separate and greater emphasis given to subjective material in dealing with psychological issues (Cape & Stiles, 1998).

DESCRIPTIONS ARE NOT EVALUATIONS

Although precise descriptions can yield insights into the processes and mechanisms of interviews, they cannot directly reveal which interviews are good ones. The problem is that skillful interviewing requires responsiveness to the emerging circumstances rather than simply increasing or decreasing the use of particular components. Human interaction displays sensitive dependence on initial conditions; that is, it is chaotic in a technical sense (Barton, 1994). A chance remark or gesture can send an interview off in a direction that could not have been predicted beforehand. For practical purposes, this problem is intractable, just as is predicting the weather far in advance. Like the weather, we can describe many general characteristics of medical interviews (warm in the summer, cold in the winter; medical histories consist mainly of closed question and exposition exchanges), but we cannot predict specific events far in advance (when it will rain, what new information will emerge).

The responsiveness problem may be obscured by occasional findings linking descriptive interview components to outcomes. For example, in our earlier work, the proportion of the medical history devoted to exposition exchanges was moderately correlated with patient satisfaction (Putnam *et al.*, 1985; Stiles *et al.*, 1979). Such findings must be interpreted very carefully. They might reflect a specific failure of responsiveness (perhaps many patients in these studies engaged in less exposition than they would have liked, so that those who did more were relatively more satisfied), but they are not a trustworthy guide to the value of different interview components. For example,

even though the proportion of closed question exchanges has generally shown null correlations with satisfaction and adherence in such studies, it would be absurd to suggest that closed questions were worthless or that physicians should ignore or avoid closed questioning. Most experts would agree that closed questions are essential for efficiently gathering information needed for diagnosis and treatment. The null correlations with outcomes probably reflected appropriate responsiveness: most interviews had about as much of this type of exchange as was required, so that outcomes did not vary as a function of whether there was a surplus or a deficit.

EVALUATIONS INCORPORATE JUDGMENTS OF APPROPRIATE RESPONSIVENESS

Evaluations of medical interviews, I suggest, incorporate judgments of whether the interviewer was appropriately responsive. The value of a closed question or other interview behavior depends on whether it was used appropriately, that is, whether it was responsive to the emerging context, patient requirements, and medical goals. There may be no strong or consistent relation between interview descriptions and evaluations. Any particular behavior could be evaluated positively if it was appropriate under prevailing circumstances and negatively if it was not. Evaluative measurements, for example, judging interviewing skill—draw on expert or intuitive knowledge and understanding of context. The measures often depend on empathic or experiential processes—understanding the patient's thoughts and feelings, for example. Although experts can make reliable evaluative judgments of interviews, and point to specific events that support their judgments, it is frustratingly difficult to say precisely how to construct a good relationship or a good interview or a even a good intervention (cf. Pirsig, 1974).

Evaluative measures are more likely than descriptive measures to show linear relations to positive outcomes. Physicians who try to understand their patients' concerns, treat them with respect, and are warm and caring, characteristics that are clearly positively valued, tend to have more satisfied, trusting, adherent patients than those who do not (e.g., Gabel *et al.*, 1993; Roberts & Aruguete, 2000; Thom & Campbell, 1997; Thom *et al.*, 1999; Williams *et al.*, 1998). Similarly, in psychotherapy research, the two strongest predictors of positive outcomes are the strength of the patient-therapist alliance (Horvath & Bedi, 2002), and therapist

empathy (Bohart *et al.*, 2002), both of which are evaluative and depend on appropriate responsiveness. A strong alliance, like any good doctor-patient relationship, is not a psychotherapeutic technique but an achievement based on the responsive use of techniques. Measuring the alliance involves judging the fit of action to context and of participants to each other. A particular type of intervention (e.g., transference interpretation, assignment of behavioral homework, closed question) may either advance or retard the alliance depending on circumstances. Appropriate responsiveness—choosing and timing interventions to fit emerging patient requirements—is incorporated into alliance ratings. Similarly, empathy in interviewing consists in responding sensitively and appropriately to the emerging experience of the patient. The particular responses that are empathic for one patient would not be empathic for another.

Scales for rating physician skill and competence also include many items that call for the rater to judge whether each activity was done well, appropriately, or properly. Raters must judge whether the behavior was appropriately responsive to the context in which it occurred. For example, the Brown Interview Checklist (Goldberg *et al.*, 1994), an evaluative rating scale designed for application to interview audiotapes, includes such items as "Calibration - Assesses the patient's ability to communicate," "IF APPROPRIATE: Interrupts to redirect," "Eye contact - Appropriate length to enhance patient comfort," "Respect - Expresses respect for the patient's coping or makes a statement of praise," "Establishes and maintains a narrative thread," and "IF APPROPRIATE: Makes appropriate follow-up arrangements." Such items serve the important purpose of systematically directing raters' attention to relevant components of the interview. However, they are not simply reports of the presence or absence (or degree) of specified behaviors but instead demand judgments about whether the physician responded appropriately to the emerging context. The same behavior might be given a high rating in one context (i.e., if it was used appropriately) and a low rating in another context (if it was inappropriate at that moment). Other evaluative measures contain similar items (e.g., Kraan *et al.*, 1989; Stewart *et al.*, 2000; Stillman *et al.*, 1977). By contrast, items in descriptive measures seek to represent behaviors independently of whether they were used appropriately or inappropriately on a particular occasion (e.g., Del Piccolo *et al.*, 2002; Stiles, 1992).

EVALUATIONS ARE NOT DESCRIPTIONS

Evaluative measures may be extremely useful for such purposes as assessing physician skills. Because they incorporate responsiveness, however, they can be disappointingly uninformative as descriptors. For example, finding that a positive doctor-patient relationship is associated with patient adherence reveals little about how to achieve adherence. Such evaluative characteristics as relationship strength, empathy, and skill are not voluntary behaviors that can be arbitrarily increased or decreased in the same way as can closed questions or information-giving. Finding associations of evaluative measures with outcomes is plausible (in general, good interviewing is associated with good outcomes) but may not provide much specific practical guidance.

Perhaps evaluative measures should be considered as a first step in unpacking responsive processes. For example, knowing that empathy predicts outcome allows one to ask, "What processes contribute to empathy?"--perhaps a slightly more tractable question than "What processes contribute to outcome?" Considering both evaluative process measures and evaluative outcome measures as manifestations of responsive dynamic processes directs attention to multiple levels of description and analysis. The unpacking process may yield more easily to alternative, qualitative approaches (Stiles, in press).

CONCLUSION

High-quality descriptive and evaluative measures of medical interview process components are available, but they are differently constructed and serve different purposes. Because human interaction is intricately responsive on many time scales, one cannot expect to find precise descriptions of what makes a good interview. Good interviewing reflects appropriate responsiveness to emerging characteristics of the interaction that cannot be predicted in advance, even though an interview's quality can be reliably assessed in retrospect. The development of separate technologies for description and evaluation of interviews reflects the responsive nature of the phenomenon rather than any inadequacy of descriptive or evaluative measures.

Acknowledgments. I thank Samuel M. Putnam, Arthur C. Bohart, and Karen Tallman for helpful discussions and comments.

REFERENCES

- Barton S. (1994). Chaos, self-organization, and psychology. *American Psychologist*, 49, 5-14.
- Bohart A.C., Elliott R., Greenberg L.S. & Watson J.C. (2002). Empathy. In *Psychotherapy Relationships that Work: Therapist Contributions and Responsiveness to Patient Needs* (ed. J. C. Norcross), pp. 89-108. Oxford University Press: New York.
- Cape J.D. & Stiles W.B. (1998). Verbal exchange structure of general practice consultations with patients presenting psychological problems. *Journal of Health Psychology* 3, 5-21.
- Del Piccolo L., Mazzi M., Saltini A. & Zimmermann Ch. (1998). Which patients talk about stressful life events and social problems to the general practitioner? *Psychological Medicine* 28, 1289-1299.
- Del Piccolo L., Saltini A., Zimmermann Ch. & Dunn G. (2000). Differences in verbal behaviors of patients with and without emotional distress during primary care consultations. *Psychological Medicine* 30, 629-643.
- Del Piccolo L., Mazzi M., Saltini A. & Zimmermann Ch. (2002). Inter and intra individual variations in physicians' verbal behavior during primary care consultations. *Social Science and Medicine* 55, 1871-1885.
- Frankel R.M. (1984). From sentence to sequence: Understanding the medical encounter through microinteractional analysis. *Discourse Processes* 7, 135-170.
- Gabel L.L., Lucas J.B. & Westbury R.C. (1993). Why do patients continue to see the same physician? *Family Practice Research Journal* 13, 133-147.
- Goldberg R.J., Novack D.H., Metheny W.P., Dube C. & Goldstein M.G. (1994). Using an interview-content-assessment instrument to evaluate a course in medical interviewing. *Academic Medicine* 69, 853-853.
- Goodwin C. (1981). *Conversational Organization: Interaction Between Speakers and Hearers*. Academic Press: New York.
- Grice H.P. (1975). Logic and conversation. In *Syntax and Semantics*, Vol. 3: *Speech Acts*. (ed. P. Cole and J.L. Morgan). Academic Press: New York.
- Horvath A.O. & Bedi R.P. (2002). The alliance. In *Psychotherapy Relationships that Work: Therapist Contributions and Responsiveness to Patient Needs* (ed. J.C. Norcross), pp. 37-69. Oxford University Press: New York.
- Kraan H., Crijnen A., Zuidweg J., van der Vleuten C. & Imboos T. (1989). Evaluating undergraduate training. In *Communicating with Medical Patients* (ed. M. Stewart and D. Roter), pp. 167-177. Sage Publications: Newbury Park, CA.
- McDaniel S.H., Stiles, W.B. & McGaughey, K J. (1981). Correlations of male college students' verbal response mode use in psychotherapy with measures of psychological disturbance and psychotherapy outcome. *Journal of Consulting and Clinical Psychology* 49, 571-582
- Miller W.R., Benefield G. & Tonigan J.S. (1993). Enhancing motivation for change in problem drinking: a controlled comparison of two therapist style. *Journal of Consulting and Clinical Psychology* 61, 455-461.
- Pirsig R.M. (1974). *Zen and the Art of Motorcycle Maintenance: an Inquiry into Values*. Morrow: New York.
- Putnam S.M. (2002). Changing health behaviors through the medical interview. *Epidemiologia e Psichiatria Sociale* 11, 218-225.
- Putnam S.M. & Stiles W.B. (1993). Verbal exchanges in medical interviews: implications and innovations. *Social Science and*

- Medicine* 36, 1597-1604.
- Putnam S.M., Stiles W.B., Jacob M.C. & James S.A. (1985). Patient exposition and physician explanation in initial medical interviews and outcomes of clinic visits. *Medical Care* 23, 74-83.
- Putnam S.M., Stiles W.B., Jacob M.C. & James S.A. (1988). Teaching the medical interview: an intervention study. *Journal of General Internal Medicine* 3, 38-47.
- Rimondini M. & Del Piccolo L. (2002). L'approccio centrato sul paziente: dalla medicina generale alla psichiatria. *Epidemiologia e Psichiatria Sociale* 11, 232-236.
- Roberts C.A. & Aruguete M.S. (2000). Task and socioemotional behaviors of physicians: a test of reciprocity and social interaction theories in analogue physician-patient encounters. *Social Science and Medicine* 50, 309-315.
- Sacks H., Schegloff E.A. & Jefferson G.A. (1974). A simplest systematics for the organization of turn-taking in dyadic conversation. *Language* 50, 697-735.
- Stewart M., Brown J.B., Donner A., McWhinney I.R., Oates J., Weston W.W. & Jordan J. (2000). The impact of patient-centered care on outcomes. *Journal of Family Practice* 49, 796-804.
- Stiles W.B. (1989). Evaluating medical interview process components: null correlations with outcomes may be misleading. *Medical Care* 27, 212-220.
- Stiles W.B. (1992). *Describing Talk: a Taxonomy of Verbal Response Modes*. Sage: Newbury Park, CA.
- Stiles W.B. (1996). Stability of the verbal exchange structure of medical consultations. *Psychology and Health* 11, 773-786.
- Stiles W.B. (in press). Qualitative research: evaluating the process and the product. In *Handbook of Clinical Health Psychology* (ed. S.P. Llewelyn and P. Kennedy). Wiley: London.
- Stiles W.B. & Putnam S.M. (1992). Verbal exchanges in medical interviews: concepts and measurement. *Social Science and Medicine* 35, 347-355.
- Stiles W.B. & Shapiro D.A. (1994). Disabuse of the drug metaphor: psychotherapy process-outcome correlations. *Journal of Consulting and Clinical Psychology* 62, 942-948.
- Stiles W.B. & Shapiro D.A. (1995). Verbal exchange structure of brief psychodynamic-interpersonal and cognitive-behavioral psychotherapy. *Journal of Consulting and Clinical Psychology* 63, 15-27.
- Stiles W.B., Putnam S.M., Wolf M.H. & James, S.A. (1979). Interaction exchange structure and patient satisfaction with medical interviews. *Medical Care* 17, 667-681.
- Stiles W.B., Honos-Webb L. & Surko M. (1998). Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice* 5, 439-458.
- Stillman P.L., Brown D.R., Redfield D.R. & Sabers D.L. (1977). Construct validation of the Arizona Clinical Interview Rating Scale. *Educational and Psychological Measurement* 37, 1031-1038.
- Thom D.H. & Campbell B. (1997). Patient-physician trust: an exploratory study. *Journal of Family Practice* 44, 169-176.
- Thom D.H., Ribisl K.M., Stewart A.L. & Luke D.A. (1999). Further validation and reliability testing of the Trust in Physician Scale. *Medical Care* 37, 510-517.
- Williams G.C. & Deci E.L. (2001). Activating patients for smoking cessation through physician autonomy support. *Medical Care* 39, 813-823.
- Williams G.C., Rodin G.C., Ryan R.M., Grolnick W.S. & Deci E.L. (1998). Autonomous regulation and long-term medication adherence in adult outpatients. *Health Psychology* 17, 269-276.