METHODOLOGY

Patient–provider communication data: linking process and outcomes in oncology care

Lisa Kennedy Sheldon[1,2]
Fangxin Hong[3,4]
Donna Berry[4,5]

[1]University of Massachusetts Boston, Boston, MA, USA; [2]St Joseph Hospital, Nashua, NH, USA; [3]Dana-Farber Cancer Institute, Department of Biostatistics and Computational Biology, Boston, MA, USA; [4]Dana-Farber Cancer Institute, Phyllis F Cantor Center for Research in Nursing and Patient Care Services, Boston, MA, USA; [5]Harvard Medical School, Boston, MA, USA

Overview: Patient–provider communication is vital to quality patient care in oncology settings and impacts health outcomes. Newer communication datasets contain patient symptom reports, real-time audiofiles of visits, coded communication data, and visit outcomes. The purpose of this paper is to: (1) review the complex communication processes during patient–provider interaction during oncology care; (2) describe methods of gathering and coding communication data; (3) suggest logical approaches to analyses; and (4) describe one new dataset that allows linking of patient symptoms and communication processes with visit outcomes.

Challenges: Patient–provider communication research is complex due to numerous issues, including human subjects' concerns, methods of data collection, numerous coding schemes, and varying analytic techniques.

Data collection and coding: Coding of communication data is determined by the research question(s) and variables of interest. Subsequent coding and timestamping the behaviors provides categorical data and determines the interval between and patterns of behaviors.

Analytic approaches: Sequential analyses move from descriptive statistics to explanatory analyses to direct analyses and conditional probabilities. In the final stage, explanatory modeling is used to predict outcomes from communication elements. Examples of patient and provider communication in the ambulatory oncology setting are provided from the new Electronic Self Report Assessment-Cancer II dataset.

Summary: More complex communication data sets provide opportunities to link elements of patient–provider communication with visit outcomes. Given more complex datasets, a step-wise approach is necessary to analyze and identify predictive variables. Sequential analyses move from descriptive results to predictive models with communication data, creating links between patient symptoms and concerns, real-time audiotaped communication, and visit outcomes. The results of these analyses will be useful in developing evidence-based interventions to enhance communication and improve psychosocial outcomes in oncology settings.

Keywords: communication, analysis, distress, cancer, outcomes

Overview

Patient–provider communication is vital to quality patient care, especially in oncology settings.[1,2] This communication impacts patient outcomes, including psychosocial adjustment and adherence to treatment recommendations.[3–5] Experts in communication research have called for more rigorous methodologies to move from descriptive studies and programs developed from expert opinion to evidence-based interventions and experimental designs.[6–8] With these methodologies, communication data can be explored to reveal the temporal and sequential relationships between patient and provider behaviors and outcomes. These relationships provide the information to create predictive

Correspondence: Lisa Kennedy Sheldon
University of Massachusetts Boston,
100 Morrissey Boulevard, Boston
MA 02125, USA
Tel +1 603 582 2845
Fax +1 617 287 7500
Email lisa.kennedysheldon@umb.edu

311

models for communication outcomes that will be superior in experimental designs. Analysis of these relationships requires a step-wise approach to identify the communication elements that effect specific patient and visit outcomes.

Earlier research in patient–provider communication has often been exploratory and descriptive.[9,10] Increasingly, leaders in the field are calling for more evidence linking provider communication with patient outcomes to demonstrate the effectiveness of specific communication approaches.[5,11–13] Identifying patterns within patient–provider communication is necessary to determine the characteristics of behaviors such as type and frequency or dose that impact desired patient outcomes. Newer datasets provide more complete information about patient–provider communication elements and visit outcomes including all three key elements: patient self-report of symptoms, audiotaped patient–provider communication during the visit, and visit outcomes. Using all three elements provides information about communication processes and their association to patient and visit outcomes.

This paper reviews challenges in data collection and processing of patient–provider communication, describes approaches to step-wise data analyses, and then provides examples from one new dataset that captures patient self-report of symptoms, real-time audiofiles of patient–provider communication, and visit outcomes.

Challenges

The study of communication between patients and health care providers is essential to understanding the processes that predict the desired patient outcomes. Issues that increase the complexity of communication research include human subjects' concerns, types of data collection, numerous coding schemes, and varying analytic techniques.[5] One concern is that the process of being observed or recorded may affect actual behavior. Additionally, there are ethical concerns with regard to privacy when these interactions are recorded. One solution to these ethical concerns is the use of actors as "simulated" (standardized) patients who interact with providers using a script.[14] However, the use of simulated patients adds an element of artificiality and lack of situational and relational context inherent in real clinical encounters. Additionally, in one study, investigators concluded that larger sample sizes were required in studies using simulated patients to increase reliability in assessment and coding of provider communication.[15]

Data collection and processing

Data collection in communication studies has been accomplished by numerous methods, including direct observation and audio- and video-recording patient–provider conversations. In addition, studies are designed to capture any number of desired variables, including specific behavior(s) from a variety of subject(s) such as real or simulated patient(s), provider(s), and/or caregiver(s). Depending on the collection method and desired individual, paired, and group variables, the data are often multivariate and complex. This complexity creates issues in coding behaviors, requiring application of appropriate and logical statistical analyses to reduce Type I errors and increase generalizability.[16–18] Additionally, data must be coded and include the capture of the temporal elements for subsequent analyses.

Coding issues

After data collection, patient and provider communication is coded by applying a selected scheme to capture the specific variables or concepts under investigation. In the past, most coding was done with the audio- or video-taped patient–provider communication; there was no context of patient perspective on symptoms or concerns or the related visit outcomes.

Communication between patients and providers may be coded in terms of overall qualities (eg, biomedical or psychosocial), the nature of behaviors (eg, instrumental or affective), provider style (eg, facilitative or dominant), the type of behavior (open- or closed-ended question) or by patient/provider demographic variables (eg, gender, patient diagnosis, or provider profession). While coding is a cumbersome and time-consuming process, coding specific patient and provider behaviors provides data for more complex correlative and predictive analyses. For example, health care providers often provide reassurance to anxious patients with a cancer diagnosis but when this behavior is unrelated to a specific health state, anxiety may actually increase in some patients.[19] Finally, issues regarding interrater reliability arise during coding that require consistent coder training, double coding of a portion of recordings, and post hoc analyses of reliability.

Numerous coding schemes have been developed for patient and provider health-related communication including the Roter Interaction Analysis System (RIAS),[17] the Medical Interview Process System,[20] the Medical Interview Aural Rating System (MIARS),[21] and the more recent Verona Coding Definitions of Emotional Sequences.[22] Most approaches code a specific behavior, turn, utterance, or cue within the communication. For example, in RIAS, an utterance is the smallest unit of expression to which a code can be assigned.[17] An utterance, in this coding scheme, contains

one complete thought and there may be multiple utterances within a sentence or recorded passage.

Observers or coders may code a behavior(s) in multiple ways. A behavior may be coded by occurrence as a "criterion utterance" or "given behavior" such as a patient cue of distress and a "target utterance" or "adjacent behavior" such as provider acknowledgment.[23] For example, in the MIARS, the target behavior is a patient cue of emotional concerns.[21] However, depending on the coding scheme and purpose of the coding, each cue or behavior may contain multiple types of commonly coded behaviors. For example, in the MIARS coding system, provider behaviors in response to patient cues of emotional concerns include acknowledgment, exploration, and/or distancing behaviors. Deciding on a single, appropriate code is required, necessitating consistent training of coders and post hoc analyses of reliability.

Temporal issues

The temporal nature of provider and patient behaviors may be captured by timestamping the recording and determining the interval between patient cue and provider behavior, or "lag." For example, did the provider facilitate the patient's disclosure of an emotional concern or, in MIARS, a cue? According to Bakeman and Quera[24] a lag may be "positive," the association between a given behavior and the immediately *following* behavior, or "negative," the association between a "given behavior" and the immediately *preceding* behavior. Additionally, behaviors may be coded by interval (behavior and time) or by event (behavior occurrence and order). Using a timestamp for behavioral events allows for the identification of temporal relationships between one behavior and adjacent behaviors and also for the timing of events within the interaction. As seen in Figure 1, the temporal nature of patient and provider communication may be seen as patterns within the interaction, patient cues, and provider responses, or provider facilitation of patient concerns through questions. Finally, understanding the patient's concerns before the actual visit through pre-assessment may provide context for the recorded communication during the visit.

New datasets: patient symptoms, audiofiles, and visit outcomes

Newer datasets, using electronic data collection methods, contain more extensive patient and provider variables and associated visit outcomes. One dataset gathered during testing of the Electronic Self Report Assessment-Cancer (ESRA-C)[27] contains patient-reported symptoms and quality of life data collected prior to the ambulatory visit and digital audiofiles

of real-time patient–provider communication recorded during the visit. Additionally, the dataset contains the visit outcomes, capturing the full spectrum of information: patient symptoms, audiotaped communication, and visit outcomes.

Given the comprehensive information available in these new datasets, a coding scheme needs to be matched to the specific aims and hypotheses. For example, the ESRA-C dataset contains patient symptoms including psychosocial concerns as well as embedded measures of depression, emotional functioning, and outlook. One coding scheme, MIARS, provides a method to code the information about patient cues of psychosocial and emotional concerns and the preceding and/or following provider behaviors. In the following example from this dataset, a male patient is talking with a male provider in an ambulatory oncology clinic 6 weeks after receiving a stem cell transplant. The patient described his feelings and the provider is trying to acknowledge and normalize the patient's feelings.

> Patient: "I'm not at my lowest, but I get a lot of the 'hopeless, what's the use feeling.'"
> Provider: "Uh, huh…"
> Patient: "Anxiety, fear, this whole stem cell killing my bone marrow…"
> Provider: "That's normal to have a hard time on that. Yeah."

In this interaction, MIARS coding can capture the patient cue (MIARS Level II – *direct expression of a concern*) and the provider's *acknowledgment* of the patient's feelings. The dataset also includes the patient's symptoms from self-report and the visit outcomes. From the patient summary report of symptoms, the patient scored moderately high on depression but the provider did not have this information. Rather than further exploration of this patient's emotional condition, the provider stopped the discussion after the acknowledgment. Further assessment may have given the provider more information to determine the level of the patient's distress and the need for further evaluation, treatment, and/or referral. Ultimately, having the patient summary report before the visit may have directed the conversation toward further exploration of the patient's psychosocial wellbeing.

In the next excerpt from another provider in the same clinic, the provider acknowledged and explored the patient concerns. The patient has expressed her concerns with audible crying (MIARS Level III cue). The provider has acknowledged her feelings and asks an exploratory question to further delve into her emotional state. Both acknowledgment and exploration are necessary components of full assessment of psychosocial concerns, particularly in oncology settings.
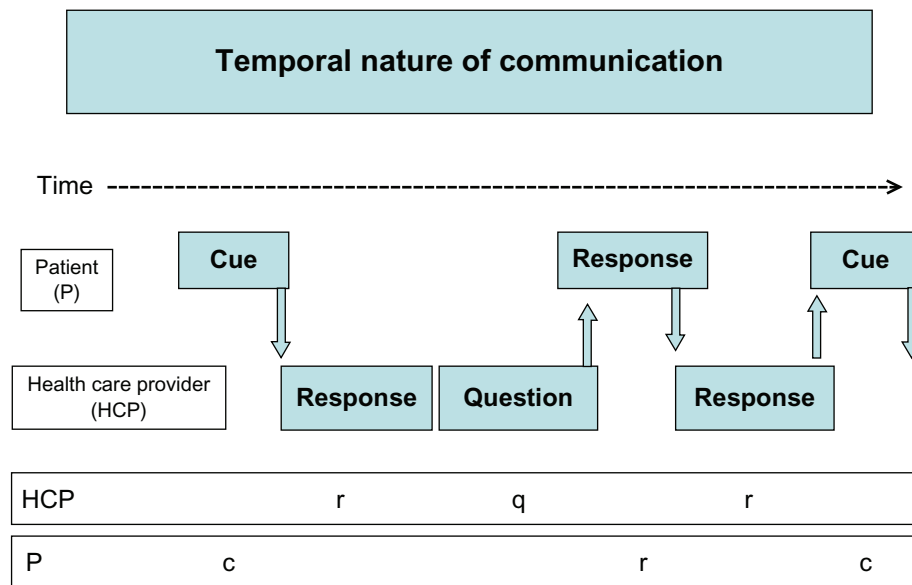
Figure 1 The temporal nature of patient–provider communication.

> Patient: "Not only that, my spiritual advisor told me that
> crying helps release anxiety that I am feeling about
> every stage [after transplant]."
> Provider: "Uh, huh. Can you still enjoy things?"

While some providers may acknowledge and explore patient concerns, others may use *distancing behaviors* to control or curtail the discussion. Distancing behaviors may include changing the topic, ignoring the topic completely, or providing a cursory response such as premature reassurance, and then moving on to another topic. The following example from the same clinic provides an example of a distancing behavior, *premature reassurance*, by the provider. The provider was unaware of the patient's symptom report of moderate depression and did not pursue the patient's concern about the future.

> Patient: "I'm gonna run out of luck, is all I think. I've had
> a bunch of these."
> Provider: "I know, I know. People get through it."

The provider's role in facilitating patient disclosure of concerns reframes coding to include an exploration of adjacent behaviors, both preceding and following patient cues. The link between the timing or temporal nature of provider and patient behaviors is needed to capture the nature of the communication and link it to an outcome. Such links may be process indicators that lead to specific outcomes. For example, when a patient expresses significant depressive symptoms, the appropriate provider responses would include acknowledgement and exploration of the patient's symptoms. Process indicators might include

assessment of patient psychosocial concerns and emotional wellbeing, either using valid and reliable tools or pre-identified questions. The desired outcome of the interaction, for a patient with significant psychosocial issues, would include pharmacologic treatment and/or referral for treatable conditions.

In the next example, the provider was facilitating patient disclosure of physical symptoms when a patient switched the topic to psychosocial issues. The provider followed the cue and explored the patient's response.

> Provider: "Besides the diarrhea, is there anything specifi-
> cally you've noticed that's new?"
> Patient: "I've always had mood swings, so I can't tell if
> they're more severe or not."
> Provider: "OK. But it's definitely something you are notic-
> ing now? Are you feeling much more depressed
> than usual?"

Coding the provider and patient behaviors both in detail and temporal sequence creates a pattern of interaction that can be associated with particular process indicators and outcomes of interest. The outcome of the visit, documented by ESRA-C coding and/or medical record data, may include pharmacologic treatment and referral for counseling. Additionally, if the provider has the patient symptom report, the discussion of psychosocial concerns may be more comprehensive, ultimately improving patient outcomes. The association between specific coded behaviors, process indicators, and visit outcomes provides information for predictive models linking behaviors with desired outcomes.

Data analytic approaches

Raw communication data are often collected, processed, and coded using recordings (audio or video) to create categorical data. Analyses of these data can result in identification of temporal communication patterns.[22,23,26,27] For example, a patient (P) expression or cue of a concern ("c") is usually followed by a health care provider (HCP) response (r) (see Figure 1). Using the MIARS coding scheme, provider responses to patient concerns may include acknowledgement (A), an open-ended question for exploration (E), and/or minimal encouragement (ME) such as "Uh huh."

On the other hand, the provider may initiate the patient expression of concerns, a preceding behavior. During an interaction or visit, behaviors occur at time points and can be "timestamped," providing time-series data or strings of behaviors. Stringing together the temporal sequence of the behaviors, "event sequences," provides clues as to how the conversation unfolds and the patterns of communication. In the realm of psychosocial concerns, research questions arising from these patterns may include: How many patient cues of concerns are needed to prompt provider exploration? Does minimal encouragement from the HCP increase the frequency of patient expression of concerns?

Step-wise analyses

Sequential analyses conducted in a step-wise manner are needed to move from descriptive results to predictive models (Figure 2). Initial analyses from coded communication data often include descriptive and summary statistics such as frequencies, percentages, and means of demographic information, and of patient and provider coded behaviors, for example, the number of patient cues of psychosocial concerns per visit, the percentage of patients prescribed pharmacologic treatment and referred for counseling, and mean consultation duration. Then, exploratory analyses can provide clues of possible relationships and associations between behaviors. Types of such analyses often include correlations, two-way contingency tables, and odds ratio. For example, the results would answer the relationship between provider facilitations and patient expressions of psychosocial concerns: is minimal encouragement likely followed by patient expression of concerns (significance of particular chain) or would more provider facilitations increase the number of patient expressions of concerns (positive correlation)?

A conditional probability (CP) examines the probability of an event occurring given another event has occurred. For example, if the health care provider asks a patient about pain, the patient will usually describe their pain in more detail. On the other hand, sometimes patients bring up topics such as sexuality that providers may want to avoid discussing. It requires discrete not continuous variables and a method to identify time periods. Using timestamps from audiofiles, discrete information about the occurrence of specific desired behaviors can be linked with visit outcomes. Conclusions may be confounded by the fact that explanatory variables may be correlated with other explanatory variables. Simple tests, such as t-tests and Fisher's exact test could explore the difference in mean or percentage between predefined groups, for example, length of visit or percentage referral for counseling between female and male providers. Such direct analyses are useful during exploratory and descriptive studies, but not for more complex models of communication processes where multiple variables may predict a single outcome (Figure 2).

More rigorous statistical analyses are informed by the results of the descriptive, correlational, and probabilistic analyses of the coded and timestamped communication data. In the final step, statistical modeling is used to abstract the relationship between specific target behaviors and potential explanatory variables from complex phenomena.[27] Logistic regression is often used with a discrete, dichotomous outcome variable such as discussed concerns/did not discuss concerns. These models explore the prediction of the probability of the occurrence of an event by explanatory variables such as patient age, gender, diagnosis, provider distancing behaviors, etc. For example, what provider behaviors are most associated with patient disclosure of concerns? Depending on the distribution of the outcome variables, other regression
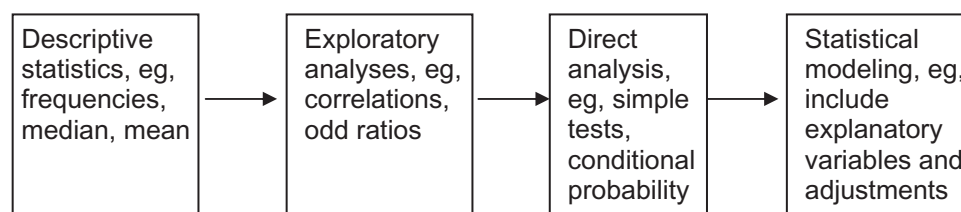
Figure 2 Sequence of analyses for coded communication data.

models may also be used including Poisson regression and log-linear approaches. For example, frequencies of cues or concerns may also be used as count data in Poisson regression models to link patient to provider behaviors.

Sequential analyses are a set of approaches to capture the complexity of patient–provider communication.[6,23,26,27] Sequential analyses treat current behaviors as probabilistically determined by recent and adjacent behaviors. These analyses attempt to explain asymmetries in probabilities with explanatory variables. Given an antecedent or preceding behavior, how does the conditional probability (CP) of a consequent or following behavior deviate from the unconditional probability (UP) of the consequent behavior? Additionally, these analyses may be useful in analyzing sequences of time-series data. For example, how often does patient disclosure of psychosocial concerns follow a provider's facilitation at the beginning of the visit as opposed to the end of the visit?

Sequential analyses require special considerations during the interpretation of the results. Considerations include temporality, contingency, redundancy, stationarity, and homogeneity. Given that interaction itself is a temporal form, the notion of temporality must be considered during the interpretation of results (Gottman and Roy, pp. 60).[28] The concept of contingency arises because behaviors in dyads are associated with adjacent behaviors.[25]

1. Redundancy – behaviors may be repeated across a visit.
2. Stationarity/Nonstationarity – stability or instability of parameters of sequential connection over time, by visit, or by source.
3. Homogeneity/Nonhomogeneity – consistency or lack thereof of sequential relationships across sources (patient/provider).

Summary

Recorded and coded patient–provider interactions provide valuable information about communication processes and visit outcomes. Capturing and coding communication data has multiple challenges including concerns about human subjects and practical concerns about coding and analyzing the communication. Newer datasets such as patient symptom self-report[25] provide valuable information about patient characteristics and outcomes. With the availability of more comprehensive electronic datasets and detailed coding, numerous variables are available to identify specific patient characteristics, patterns of behavior, and visit outcomes. Identification of desired outcomes is required to link communication patterns with outcomes and identify behaviors that

predict patient outcomes and can be integrated into clinical care.[29] Application of a step-wise approach to analyses of coded communication data is necessary to reveal generalizable results. This paper describes a step-wise approach to statistical analyses of patient and provider communication that is useful in identifying predictive variables.

Incorporating time into the analysis of coded communication data also allows researchers to capture communication as it unfolds during a visit and create communication sequences. Moving beyond the occurrence of behaviors to the actual temporal sequence of communication will be useful in predicting the timing and dosing of particular provider behaviors to achieve desired outcomes. Temporal information further informs what is known about how and when providers discuss concerns with patients, especially important in regard to patients with psychosocial concerns and diagnosable conditions such as depression and/or at risk for suicide. Results from step-wise analyses of more comprehensive datasets will provide important information about the relationships between provider and patient behaviors, identify patterns within interactions, and predict the most effective behaviors to achieve desired visit outcomes, ultimately improving patient outcomes.

Ultimately, understanding communication between providers and patients in oncology care is better informed by newer datasets. Rather than coding one visit in isolation of other patient and visit information, newer datasets provide the patient perspective on symptoms and quality of life issues. In addition to the real-time communication captured by audio- and video-taping, they also include the visit outcomes. Now, in line with current research priorities, step-wise analyses can be used to link communication elements with patient symptoms and visit outcomes. This marks the advent of a new era in studying patient–provider communication that informs the development of evidence-based interventions to improve communication and patient outcomes.

Disclosure

References

1. Epstein RM, Street RL. *Patient-Centered Communication in Cancer Care: Promoting Healing and Reducing Suffering*. Bethesda, MD: National Cancer Institute; 2007. [NIH Publication No.07-6225].

2. Adler NE, Page AEK. *Cancer Care for the Whole Patient: Meeting Psychosocial Health Needs*. Washington, DC: National Academies Press; 2008.

3. Fellowes D, Wilkinson S, Moore P. Communication skills training for healthcare professionals working with cancer patients, their families and/or careers. *Cochrane Database Syst Rev*. 2004;(2):CD003751.

4. Thorne SE, Bultz BD, Baile WF, SCRN Communication Team (2005). Is there a cost to poor communication in cancer care?: a critical review of the literature. *Psychooncol*. 2005;14:875–884.

5. de Haes H, Bensing J. Endpoints in medical communication research: Proposing a framework of functions and outcomes. *Patient Educ Couns*. 2010;74:287–294.

6. Eide H, Quera V, Finset A. Exploring rare patient behavior with sequential analysis. *Epidemiol Psichaitr Soc*. 2003;12:109–114.

7. Connor M, Fletcher I, Salmon P. The analysis of verbal interaction sequences in dyadic clinical communication: a review of methods. *Patient Educ Couns*. 2009;75:169–177.

8. Bensing JM, Verheul W, Jansen J, Langewitz WA. Looking for trouble: the added value of sequence analysis in finding evidence for the role of physicians in patients' disclosure of cues and concerns. *Med Care*. 2010;48:583–588.

9. Ford S, Fallowfield L, Lewis S. Doctor-patient interactions in oncology. *Eur J Cancer*. 1996;31:2264–2269.

10. Razavi D, Delvaux N, Marchal S, De Cock M, Farvacques C, Slachmuylder JL. Testing health care professionals' communication skills: the usefulness of highly emotional standardized role-playing sessions with simulator. *Psychooncol*. 2000;9:293–302.

11. Arora, NK. Interacting with cancer patients: The significance of physicians' communication behavior. *Soc Sci Med*. 2003;57:791–806.

12. Arora NK, Weaver KE, Clayman ML, Oakley-Girvan I, Potosky AL. Physicians' decision-making style and psychosocial outcomes among cancer survivors. *Patient Educ Couns*. 2009;77:404–412.

13. Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician–patient communication to health outcomes. *Patient Educ Couns*. 2009;74:295–301.

14. Kruijver IPM, Kerkstra A, Kerssens JJ, Holtkamp CCM, Bensing JM, van de Wiel HBM. Communication between nurses and simulated patients with cancer: evaluation of a communication training programme. *Eur J Oncol Nurs*. 2001;5:140–150.

15. Reinders ME, Blankenstein AH, Marwijk HWJ, et al. Reliability of consultation skills assessments using standardized versus real patients. *Med Educ*. 2001;45:578–584.

16. Booth K, Maguire P, Hillier VF. Measurement of communication skills in cancer care: myth or reality? *J Adv Nurs*. 1999;30:1073–1079.

17. Roter, DL. *The Roter Method of Interaction Process Analysis*. Baltimore, MD: Johns Hopkins University; 2005.

18. Carlson LE, Feldman-Stewart D, Tishelman C, Brundage, MD. Patient-professional communication research in cancer: An integrative review of research methods in the context of a conceptual framework. *Psychooncol*. 2005;14:812–828.

19. Stark D, Kiely M, Smith A, et al. Reassurance and the anxious cancer patient. *Brit J Cancer*. 2004;91:893–899.

20. Ford S, Hall A, Ratcliffe D, Fallowfield L. The Medical Interaction Process System (MIPS): An instrument for analysing interviews of oncologists and patients with cancer. *Soc Sci Med*. 2000;50:553–566.

21. Heaven CM, Green C. *Medical Interview Aural Rating Scale*. Psychological Medical Group. Manchester, UK: Christie Hospital, Stanley House; 2001.

22. Del Piccolo L, de Haes H, Heaven C, et al. Development of the Verona coding definitions of emotional sequences to code health providers' responses (VR-CoDES-P) to patient cues and concerns. *Patient Educ Couns*. 2011;82(2):149–155.

23. Bakeman R, Quera V, Gnisci A. Observer agreement for timed-event sequential data: A comparison of time-based and event-based algorithms. *Behav Res Methods*. 2009;41:137–147.

24. Bakeman R, Quera V. Analyzing interaction: sequential analysis with SIDS & QSEQ. Cambridge, UK: Cambridge University Press; 1995.

25. Berry DL, Blumenstein BA, Halpenny B, Wolpin, et al. Enhancing patient-provider communication with the electronic self-report assessment for cancer: A randomized trial. *J Clin Oncol*. 2011;29: 1029–1035.

26. Bakeman R, Gottman JM. *Observing interaction*: *An introduction to sequential analysis*. 2nd ed. Cambridge, UK: Cambridge University Press; 1997:60.

27. Eide H, Quera V, Guargaard P, Finset A. A physician-patient dialogue surrounding patient expression of concern: applying sequence analysis to RIAS. *Soc Sci Med*. 2004;59:145–155.

28. Gottman JM, Roy AK. *Sequential Analysis: A Guide for Behavioral Researchers*. Cambridge, UK: Cambridge University Press; 1990.

29. Berry DL. Patient-reported symptoms and quality of life integrated into clinical cancer care. *Sem Oncol Nurs*. 2011. In press.