The Physician Recommendation Coding System (PhyReCS): A Reliable and Valid Method to Quantify the Strength of Physician Recommendations During Clinical Encounters

Karen A. Scherr, BS, Angela Fagerlin, PhD, Lillie D. Williamson, BS, J. Kelly Davis, BA, Ilona Fridman, MPhil, Natalie Atyeo, Peter A. Ubel, MD

Background. *Physicians' recommendations affect patients' treatment choices. However, most research relies on physicians' or patients' retrospective reports of recommendations, which offer a limited perspective and have limitations such as recall bias.* ***Objective.*** *To develop a reliable and valid method to measure the strength of physician recommendations using direct observation of clinical encounters.* ***Methods.*** *Clinical encounters (n = 257) were recorded as part of a larger study of prostate cancer decision making. We used an iterative process to create the 5-point Physician Recommendation Coding System (PhyReCS). To determine reliability, research assistants double-coded 50 transcripts. To establish content validity, we used 1-way analyses of variance to determine whether relative treatment recommendation scores differed as a function of which treatment patients received. To establish concurrent validity, we examined whether patients' perceived treatment recommendations matched our coded recommendations.* ***Results.*** *The PhyReCS was highly reliable (Krippendorf's alpha = 0.89, 95% CI [0.86, 0.91]). The average relative treatment recommendation score for each treatment was higher for individuals who received that particular treatment. For example, the average relative surgery recommendation score was higher for individuals who received surgery versus radiation (mean difference = 0.98, SE = 0.18, $P < 0.001$) or active surveillance (mean difference = 1.10, SE = 0.14, $P < 0.001$). Patients' perceived recommendations matched coded recommendations 81% of the time.* ***Conclusion.*** *The PhyReCS is a reliable and valid way to capture the strength of physician recommendations. We believe that the PhyReCS would be helpful for other researchers who wish to study physician recommendations, an important part of patient decision making.* ***Key words:*** *prostate cancer, qualitative methods, physician-patient communication, shared decision making.* ***(Med Decis Making 2017;37:46–55)***

There has been an increasing interest in empowering patients as informed consumers of health care goods and services.[1] As informed consumers, patients often must choose between multiple treatment options. For example, in early-stage prostate cancer, patients must choose whether to receive surgery, radiation, or active surveillance. Each of these treatment options is associated with a unique profile of risks and benefits, and therefore, there is not a single right treatment option for all patients.[2] Shared decision making is considered by many to be the "pinnacle of patient-centered care," a process by which patients and physicians work together to choose the best treatment based on both medical factors and patients' individual preferences.[3] As part of this process, physicians may provide patients with recommendations. It is vital to be able to accurately capture these recommendations. Even within the paradigm of shared decision making, physicians' recommendations strongly affect patients' treatment choices, potentially even more so than patients' cancer severity, age, and anxiety.[4–6]

However, current research on physician recommendations has several limitations. Physician recommendations are frequently treated as binary, in which a physician either does or does not recommend a single treatment option.[4] In reality, however, treatment recommendations are often more nuanced, and physicians can provide recommendations of

DOI: 10.1177/0272989X16654692

varying strength for multiple treatments. In addition, most studies have relied on patient reports of physician recommendations, which are subject to recall bias.[7,8] Furthermore, motivated cognition may lead patients to misremember physician recommendations, such that their reported recommendation matches their treatment choice rather than accurately reflecting their conversation with the physician.[9]

To address these limitations, we developed the Physician Recommendation Coding System (PhyReCS), which captures the strength of physician recommendations during appointments within the context of early-stage prostate cancer. The PhyReCS addresses the aforementioned limitations in the following ways: it is a continuous (rather than binary) measure, has the flexibility to capture multiple nuanced recommendations, and avoids problems associated with relying on patients' retrospective reports of recommendations. In this article, we provide an in-depth explanation of the PhyReCS, measure its reliability, and assess its validity.

METHODS

Setting and Study Population

Appointments (n = 257) were recorded and transcribed as part of a larger trial in which men undergoing prostate biopsies were randomized to receive either a standard or low-literacy prostate cancer treatment decision aid prior to choosing a treatment for their early-stage prostate cancer.[10] The type of decision aid did not influence our measures of interest; therefore, it is not discussed further in this article. (The type of decision aid did not affect the following variables: patient treatment choice, $\chi^2[2] = 2.93$, $P = 0.23$; physician recommendation scores [e.g., for active surveillance, $F[1, 252] = 0.16$, $P = 0.69$] and patients' perceived recommendations, $\chi^2[3] = 2.29$, $P = 0.51$.) Appointments were recorded from 2008 to 2012 at 4 geographically dispersed, academically affiliated Veterans Affairs medical centers. During each appointment, the patient and physician discussed treatment options for the patient's newly diagnosed early-stage (low or intermediate risk) prostate cancer. Patient and physician demographics are listed in Table 1. There were 47 unique physicians in our study, most of whom were residents or fellows. On average, each physician was recorded in 5.31 clinical encounters (s = 3.77).

Table 1 Patient and Physician Demographics

Characteristic	Patients (n = 257)[a]	Physicians (n = 47)[b]
Age ($\bar{x}$, s)	63.2 (6.02)	33.2 (5.6)
Gender		
Male (%)	100	80
Race (%)		
White	70	68
Black	26	7
Other[c]	4	25
Education (%)		
High school or less	30	
Some college, no degree	36	
Associate's or greater	34	
Cancer risk level (%)		
Low[d]	46	

a. All demographic information is missing for 3 patients. Gleason score is missing for 2 additional patients.
b. All demographic information missing for 5 physicians.
c. "Other" category includes patients and physicians who identify as Asian, Hispanic, Native American, and multiracial.
d. All other patients had intermediate risk cancer (by study design).

Scale Development

We identified a subset of transcripts using maximum variation purposeful sampling techniques, in which we selected transcripts that differed on variables that we expected to influence physician recommendations (e.g., age, Gleason Score).[11] We then used an iterative process to develop a 5-point

Received 18 June 2015 from the Fuqua School of Business, Duke University, Durham, NC (KAS, LDW, JKD, PAU); School of Medicine, Duke University, Durham, NC (KAS, PAU); Departments of Internal Medicine and Psychology, Center for Bioethics and Sciences in Medicine, University of Michigan Ann Arbor (AF); The Ann Arbor VA Center for Clinical Management Research, Ann Arbor, MI (AF); Columbia Business School, Columbia University, New York, NY (IF); Duke University, Durham, NC (NA); Sanford School of Public Policy, Duke University, Durham, NC (PAU). Financial support for this study was provided in part by the following institutions: IIR Merit Award from the US Department of Veterans Affairs (IIR 05-283) to Angela Fagerlin, a Health Policy Investigator Award from the Robert Wood Johnson foundation to Peter A. Ubel, and Federal grant T32 GM007171 to Karen A. Scherr as part of the Medical Scientist Training Program. All funding agreements ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. Peter A. Ubel is a consultant for Humana. The principal investigator and all other coauthors have no conflicts of interest. Revision accepted for publication 6 May 2016.

Supplementary material for this article is available on the *Medical Decision Making* Web site at http://mdm.sagepub.com/supplemental.

Address correspondence to Karen A. Scherr, Fuqua School of Business, 100 Fuqua Dr., Duke University, Durham, NC 27708, USA; telephone: (281) 380-9948; e-mail: karen.scherr@duke.edu.

Physician Recommendation Scale to capture how physicians portrayed each treatment option during the clinical appointment as a whole. We defined the boundaries of each recommendation score through repeated application and discussion.

For each treatment option (surgery, radiation, and active surveillance), recommendations were coded as follows: +2 (–2) indicated that the physician made a strong recommendation for (against) the treatment, +1 (–1) indicated that the physician made a mild recommendation for (against) the treatment, and 0 indicated that the physician recommended neither for nor against the treatment. (We coded brachytherapy recommendations separately from external beam. However, it was discussed in only 35% [90/257] of appointments and received the highest recommendation score in only 5 appointments. ''Radiation'' recommendation scores thus reflect external beam radiation recommendation scores.) If the physician did not mention a treatment, this was coded as ''not discussed''; for the purpose of these analyses, this was treated as equivalent to a strong recommendation against the treatment option (–2), because such an omission essentially indicated that the physician did not think it was even worth mentioning the option. Such omissions occurred relatively infrequently (surgery: $n = 3$; radiation: $n = 3$; active surveillance: $n = 11$). Thus, for each appointment, coders assigned a recommendation score for each of the 3 primary treatment options (surgery, radiation, and active surveillance). Importantly, recommendation scores were independent such that a recommendation against a particular treatment did not automatically translate into a recommendation for another treatment.

Although recommendation scores were global judgments that considered the appointment in its entirety, there were often key statements that captured the sentiment of physicians' feelings toward a particular treatment option. Table 2 provides examples of these types of statements. As noted above, however, keep in mind that final recommendation scores were global scores based on the entire appointment rather than any single statement in isolation; therefore, we provide an example of how recommendation scores evolved over the course of an appointment in Table 3.

Coder Training and Scale Application

The lead researcher trained 5 research assistants (RAs) using the finalized codebook (available in the online appendix): each RA received approximately 30 h of guided practice over a period of 2 to 3 wk until they demonstrated a thorough understanding of the PhyReCS. RAs then double coded a random subset of 50 previously unseen transcripts, which we used to calculate reliability. Discrepancies were resolved via team discussion. Given the high reliability (see below), it was appropriate for RAs to single code the remainder of encounters ($n = 207$). All coding was finished within 3 wk, minimizing the possibility of coder drift. RAs coded the transcripts in the development set at the end of the coding period to minimize the chance of carryover from the training period. Example transcripts are available upon request.

To determine reliability, we calculated Krippendorf's alpha for the recommendation scores for the subset of transcripts that were double coded ($n = 50$). Krippendorf's alpha offers advantages over other ratings of interrater reliability such as interclass correlation coefficient or weighted kappa, and it can be used ''regardless of the number of observers, levels of measurement, sample sizes, and presence or absence of missing data.''[12,13] We treated the scale as an interval variable and used Hayes' 2013 SPSS macro to calculate Krippendorf's alpha using 5000 bias-corrected bootstrap samples to determine 95% confidence intervals.[12] (We also calculated Krippendorf's alpha treating the scores as ordinal variables and results were substantively similar.) A Krippendorf's alpha value of 0 indicates that the reliability was no better than would be expected by chance, whereas a value of 1 indicates perfect reliability; values of 0.67 to 0.8 are considered indicative of acceptable reliability, and values greater than 0.8 indicate excellent reliability.[13]

Treatment Received

Patient treatment choice was determined via chart review 6 mo after the recorded appointment (data available for 216 individuals). (Because only 5 patients received brachytherapy, we collapsed external beam radiation and brachytherapy into a single ''radiation'' category. Results remain substantively similar if we conduct analyses of brachytherapy and external beam therapy as separate categories.) Five individuals received a treatment other than active surveillance, radiation, or surgery (e.g., hormone therapy); these individuals were not included in analyses that examined which treatment patients received. Therefore, for analyses involving patient treatment choice, $n = 211$.

Table 2 For Each Recommendation Score, Examples of Key Statements That Capture the General Sentiment of Physicians' Feelings Towards Each Treatment Option

Recommendation Score	Treatment	Exemplar
−2	Active surveillance	• You're a Gleason 7. What does that mean? We shouldn't just sit tight on this because it can cause problems later in life.
	Radiation	• So I chatted with my boss and he says with [your] urinary symptoms, he agrees that radiation is probably not going to be an option that we should keep on the table.
	Surgery	• People [like you] who have a lot of tissue around their abdomen, [that] makes it dangerous to do surgery. We don't want to put you through surgery if you are at high risk, and you would be at high risk for surgery.
−1	Active surveillance	• You also have the option to [do] what's called active surveillance. . . . The fact of the matter is you had a few biopsies that were positive, and most people would say that you should do something about it rather than just watching it, although that still is an option for you.
	Radiation	• Most people, given your young age, will probably not recommend the radiation therapy. However, if you want to hear more about it, I am more than happy to ask for a counseling session with the radiation guys.
	Surgery	• I think any of them are reasonable. I think probably most people in your situation and age group would probably go for either the radiation treatment or the active surveillance rather than surgery.
0	Active surveillance	• I think active surveillance is a very reasonable approach . . . you have very low volume disease. You have low risk disease . . . So I do think that's a reasonable option for you.
	Radiation	• I think based on the biopsies, your PSA [prostate-specific antigen] and everything, both [surgery and radiation] are very equivalent treatment options for the prostate cancer that you have. A lot of it will come down to what your preference will be.
	Surgery	• So surgery . . . if that's something you want to shoot for, that would be fine.
1	Active surveillance	• So we'll have you see the radiation doctors, but with sort of the low-risk cancer and with the lung issues and stuff, watching it might be the best thing.
	Radiation	• So [radiation and surgery] are your 2 options . . . you're a bigger guy and so to get to your prostate [with surgery] will be more difficult. Not to say it's impossible. . . . Frequently it's a combination of that plus people are a little bit uneasy with surgery, and those people tend to go a little bit towards radiation.
	Surgery	• The reason I like surgery in a younger person is that no matter what you try, there's some risk that both [surgery and radiation] therapies will fail at some point in the future. . . . If we fail after surgery, we can radiate you. If you fail after radiation, we have fewer options.
2	Active surveillance	• Most people, me included, would say you're going to be fine. This prostate cancer has a very low likelihood of ever affecting you. We're going to watch you carefully, but we're not going to actually treat you per se for the cancer. That would be my opinion.
	Radiation	• I'll be honest with you, at your age, with your heart and everything, surgery would be a higher risk . . . I would recommend radiation. . . . I think that radiation's probably the way to go.
	Surgery	• My bias is that for somebody young and healthy, that I think would tolerate the surgery, and that I think would be around to experience some of those late side effects of radiation, I'd generally recommend surgery.

Table 3 Example of How Physician Recommendation Scores Evolved over the Course of an Appointment in Response to Specific Physician Comments

		Evolving Recommendation Score		
Physician Comment	**Coder Interpretation**	**Active Surveillance**	**Surgery**	**Radiation**
You have every option available to you in terms of how you want to proceed and my job now is to tell you what your options are and give you information. Every option has its risks; every option has its benefits.	All options are presented as neutral options.	0	0	0
You're not a perfect candidate for active surveillance, meaning we're a little bit more concerned about your disease. The people who we like to put on active surveillance are the people who we don't think are going to progress in their cancer . . . that may be the case for you, although given that you have a lot of cancer in each one of those biopsies, I'm a little bit more concerned about it. I don't think it's completely unreasonable, but it wouldn't be my first choice.	Mild recommendation against active surveillance.	−1	0	0
I think that in terms of cure, [surgery and radiation] are roughly equivalent, they're different techniques, so it's hard to compare one to another and they've never been compared directly head to head where we give half the people radiation and half the people surgery.	Surgery and radiation remain neutral options.	−1	0	0
I'm a biased person . . . the reason I like surgery in a younger person is that no matter what you try, there's risk that both therapies will fail at some point in the future. . . . If we fail after surgery, we can radiate you. If you fail after radiation, we have fewer options, you can't operate after radiation, that's why I tend to reserve radiation for older people.	Mild recommendation for surgery. Radiation remains neutral option.	−1	+1	0
I would recommend some form of treatment in the next 6 months. I wouldn't wait too much longer than that, just because I don't know how this cancer's going to behave . . . we felt a nodule on your rectal exam, it's Gleason six . . . and I'm skeptical that if I took your prostate out, I would find that it's higher grade than that.	Recommendation against active surveillance increases to strong (rather than mild).	−2	+1	0
	Final recommendation scores	−2	+1	0

Note: The "Evolving Recommendation Score" column reflects how the coder modified each recommendation score in response to the comment.

To examine the construct validity of the PhyReCS, we tested whether recommendation scores differed as a function of treatment received. Given that physicians' recommendations are a strong determinant of patients' treatment choices,[4] we should find that, on average, physicians' coded recommendation scores should be higher for the treatment the patient received versus the nonchosen treatment options. We calculated a relative recommendation score for each treatment option, which was equal to that treatment recommendation score minus the average of the recommendation scores for the other 2

treatments. For example, Active Surveillance $_{relative}$ = [Active Surveillance $_{raw}$] − [(Surgery $_{raw}$ + Radiation $_{raw}$)/2]. We then used a series of 3 one-way analyses of variance (ANOVAs) to examine whether the average relative recommendation score for each treatment differed as a function of treatment received. (We conducted a number of sensitivity analyses to examine the relationship between recommendation scores and treatment received, including using raw recommendation scores and transforming recommendation scores into a single categorical variable. Results were substantively similar.) For example, we tested if the average relative surgery recommendation score differed for individuals who received surgery versus radiation versus active surveillance. If Levene's test indicated that we violated the assumption of homogeneity of variance, we used a Brown-Forsythe correction for the omnibus *F*-test and a Tamhane test for pairwise comparisons. Otherwise, we used a Bonferroni correction for pairwise comparisons.

Perceived versus Coded Recommendations

Patients' perceptions of their physicians' recommendations were determined via phone interview conducted by a professional survey company approximately 7 to 10 d after the recorded appointment (data available for 205 patients). Patients were asked, ''Did your physician provide a recommendation?'' If they indicated yes, they were then asked, ''What was the recommendation?'' with the answer choices of surgery, external beam radiation, brachytherapy, watchful waiting/active surveillance, and other. (Only 5 patients perceived that their physicians recommended brachytherapy; therefore, we collapsed external beam radiation and brachytherapy into a single ''radiation'' category. Results remain substantively similar if we treat brachytherapy and external beam therapy as separate categories.) Patients who answered ''other'' (n = 2) were excluded from analysis; thus, for analyses involving patients' perceived recommendations, n = 203.

To examine the concurrent validity of the PhyReCS, we examined the concordance between physicians' recommendations as perceived by patients (''perceived recommendations'') and physicians' recommendations as determined by coders using the PhyReCS (''coded recommendations''). Given that both patients and coders are experiencing the same conversation, if the PhyReCS recommendation scores are valid, there should be relatively high concordance between patients' perceived recommendations and our coded recommendations. However, given that patients' perceptions may be influenced by factors other than the objective occurrences during the appointment, we would not be surprised to see some differences between patients' perceptions and our coded recommendations.

We classified the perceived versus coded recommendation as a ''match'' if the treatment that the patient perceived as recommended also received the highest PhyReCS recommendation score. On the other hand, we classified the perceived versus coded recommendation as a ''mismatch'' if the treatment that the patient perceived as recommended did not receive the highest PhyReCS recommendation score. For patients who perceived that the physician did not provide a recommendation, we classified the perceived versus coded recommendation as a ''match'' if more than one treatment received the highest recommendation score and as a ''mismatch'' in all other cases.

Funding

All funding agreements ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Human Subjects Approval

This study was approved by the Institutional Review Boards at each of the participating sites; written informed consent was obtained from all patients and physicians. Additional consent to be recorded was obtained from all individuals in the clinical appointment.

RESULTS

Reliability

Each of the 50 double-coded transcripts included 3 recommendation scores (1 for surgery, radiation, and active surveillance); thus, we had 150 recommendation scores to assess the reliability of our scoring system. The Krippendorf's alpha for all treatments was 0.89 (95% CI [0.86, 0.91]), indicating excellent reliability. The Krippendorf's alpha for the individual treatments was as follows: active surveillance = 0.94 (95% CI [0.92, 0.95]), surgery = 0.87 (95% CI [0.82, 0.91]), and radiation = 0.64 (95% CI [0.49, 0.80]), all of which indicate acceptable to excellent reliability. Of note, the lower reliability

for the radiation scores was due to less variation in the scores; the Krippendorf's alpha formula takes this into account and therefore reliability goes down more with each individual discrepancy because there is a higher likelihood that matches occurred by chance. Table 4 displays the observed versus expected coincidence matrices for the scores given by the 2 RAs coding each encounter. There were 29 total discrepancies out of 150 scores assigned. One coder was responsible for 15 of these discrepancies; she received remedial training before being allowed to continue with coding. Four of the discrepancies were significant, in which coders assigned scores that were 2 points apart on the scale. Three of these discrepancies were due to a misunderstanding of the coding rules by the coder who received remedial training. The remaining 25 discrepancies involved minor disagreements, in which coders assigned scores that differed by only 1 point on the scale. Importantly, coders never disagreed about whether a physician recommended for versus against a treatment option; in other words, there were no discrepancies that involved a negative versus positive score.

Table 4 For Transcripts in the Test Set (n = 50), Observed versus Expected Coincidence Matrices for Recommendation Scores (n = 150) Assigned by Coder 1 versus Coder 2[a] on the Same Transcript

Observed coincidence matrix

		Coder 2				
		−2	−1	0	1	2
Coder 1	−2	46	7	3	0	0
	−1		10	2	0	0
	0			158	9	1
	1				10	7
	2					18

Expected coincidence matrix

		Coder 2				
		−2	−1	0	1	2
Coder 1	−2	10	4	32	5	5
	−1		1	11	2	2
	0			100	15	15
	1				2	2
	2					2

Note: Frequencies on the diagonals represent perfect matches.
a. Two coders were randomly selected from our larger pool of 5 coders to double code each transcript. Frequencies on the diagonal represent perfect matches.

Relative Recommendation Scores Differ as a Function of Treatment Received

A series of 3 separate ANOVAs revealed that, for all treatments, the relative treatment recommendation score was higher for individuals who received that treatment versus the other 2 treatments (Figure 1). The relative surgery recommendation score differed as a function of treatment received, $F(2, 208) = 53.01$, $P < 0.001$. Specifically, the relative surgery recommendation score was higher for individuals who received surgery versus radiation ($M_{\text{surgery}} = 1.69$, $SE = 0.14$ vs. $M_{\text{radiation}} = 0.52$, $SE = 0.20$; mean difference = 1.17, $SE = 0.24$, $P < 0.001$) and surgery versus active surveillance ($M_{\text{surgery}} = 1.69$, $SE = 0.14$ vs. $M_{\text{active surveillance}} = -0.27$, $SE = 0.13$; mean difference = 1.96, $SE = 0.19$, $P < 0.001$). The relative radiation recommendation score also differed as a function of treatment received, Brown-Forsythe, $F(2, 85.78) = 23.22$, $P < 0.001$. Specifically, the relative radiation recommendation score was higher for individuals who received radiation versus surgery ($M_{\text{radiation}} = 1.08$, $SE = 0.22$ vs. $M_{\text{surgery}} = 0.30$, $SE = 0.08$; mean difference = 0.78, $SE = 0.23$, $P = 0.004$) and radiation versus active surveillance ($M_{\text{radiation}} = 1.08$, $SE = 0.22$ vs. $M_{\text{active surveillance}} = -0.25$, $SE = 0.09$; mean difference = 1.33, $SE = 0.23$, $P < 0.001$). Finally, the relative active surveillance recommendation score differed as a function of treatment received, Brown-Forsythe, $F(2, 187.83) = 105.81$, $P < 0.001$. Specifically, the relative active surveillance recommendation score was higher for individuals who received active surveillance versus surgery ($M_{\text{active surveillance}} = 0.52$, $SE = 0.14$ vs. $M_{\text{surgery}} = -1.99$, $SE = 0.15$; mean difference = 2.51, $SE = 0.19$, $P < 0.001$) and active surveillance versus radiation ($M_{\text{active surveillance}} = 0.52$, $SE = 0.14$ vs. $M_{\text{radiation}} = -1.61$, $SE = 0.19$; mean difference = 2.12, $SE = 0.22$, $P < 0.001$).

Comparison of Perceived versus Coded Recommendations

There was a high level of concordance between patients' perceived recommendations and the coded recommendations determined using the PhyReCS. Overall, the perceived and coded recommendation matched in 81% (164/203) of cases. Patients perceived that their physicians recommended active surveillance in 52 appointments; the perceived and coded recommendation matched in 80% (42/52) of these appointments. Patients perceived that their physicians recommended surgery

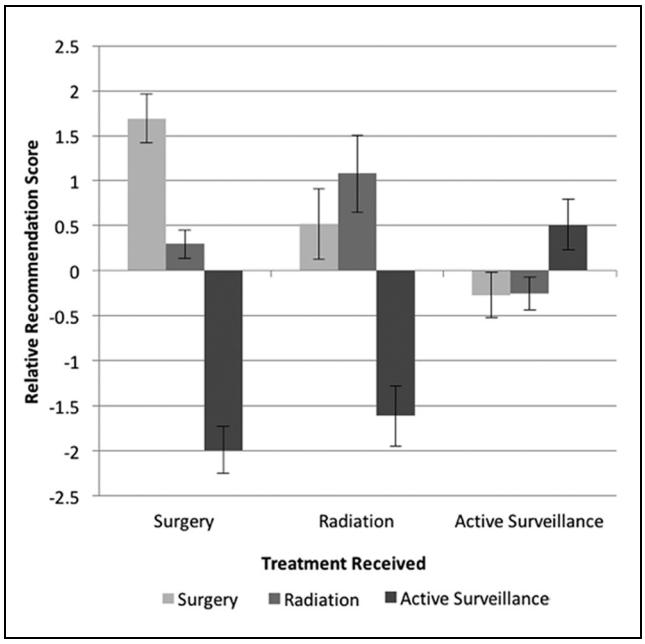

Figure 1 Average relative recommendation score for each treatment as a function of treatment received. For each treatment, the average recommendation score for the treatment received was higher than the average recommendation score for the other 2 treatments. Error bars represent ±2 SEM.

in 70 appointments; the perceived and coded recommendation matched in 91% (64/70) of these appointments. Patients perceived that their physicians recommended radiation in 26 appointments; the perceived and coded recommendation matched in 85% (22/26) of these appointments. Patients perceived that their physicians provided no recommendation in 55 appointments; the perceived and coded recommendation matched in 65% (36/55) of these appointments. The perceived and coded recommendation mismatched in 19% of cases (39/203). In 44% of these cases (17/39), patients perceived no recommendation, but the PhyReCS determined that the physician did recommend a particular treatment. Notably, the patient received the PhyReCS-recommended treatment in 65% of these cases (11/17).

DISCUSSION

In this article, we demonstrated that the PhyReCS is a reliable and valid way to quantify the strength of physician recommendations during clinical appointments in the context of early-stage prostate cancer. We showed that the scale could be applied with high reliability. We established construct validity by showing that the average relative recommendation score for each treatment was higher for individuals who received that treatment versus the other 2 treatment options. In addition, we demonstrated concurrent validity by showing that there was a high level of concordance (81%) between patients' perceived recommendations and coded recommendations as determined by the PhyReCS.

We believe that the PhyReCS would be helpful for other researchers who wish to study physician recommendations, and it is flexible enough to be adapted to many clinical settings. For example, patients with early-stage breast cancer must choose whether to receive breast-conserving therapy (lumpectomy plus radiation) or mastectomy. Like early-stage prostate cancer, the ''right'' treatment choice depends on patient preference in addition to medical factors.[14] Patient-physician conversations about these treatment options are complex, and the physician may recommend multiple treatment options with varying strength. There are, of course, important differences between these clinical situations, including the gender of patients; however, we believe that with proper validation, the PhyReCS could help to better understand the connection between physician recommendations, patient preference, and treatment choice in clinical settings besides prostate cancer.

Given the centrality of physicians' recommendations in the medical decision-making process, the PhyReCS will also allow researchers to answer other interesting and important questions. For example, future research could examine when and why there is discordance between coded treatment recommendations and patients' perceived recommendations, potentially providing insights into cognitive processes such as motivated cognition and recall bias. Are there circumstances in which patients are more (vs. less) motivated to perceive treatment recommendations as consistent with their chosen treatment option? In addition, the fact that patients often received the PhyReCS-recommended treatment when they perceived no recommendation suggests that the PhyReCS may be able to capture subtle recommendations that patients do not perceive, although future research is clearly needed to more fully examine this possibility.

The PhyReCS could also help researchers examine whether patients are more versus less satisfied with their decisions and/or clinical appointments as a function of the strength of physicians' recommendations. Patient satisfaction evaluations primarily reflect patients' perceptions of communication

with their health care providers[15–17]; therefore, it is reasonable to assume that patient satisfaction measures will be affected by differences in physician recommendations, which could be captured with the PhyReCS. It is possible that advice may decrease decision satisfaction as people like to feel that they are experiencing free choice,[18] and receiving advice can feel like an infringement on this sense of free choice.[19] Alternatively, advice may increase decision satisfaction as advice can be an important aspect of coping and patients may feel that their physicians are emotionally supportive when they give recomendations.[20] Given that recommendations change patients' sense of responsibility for a decision and responsibility is known to affect decision satisfaction,[21] the PhyReCS could also be used to examine the connection between decision responsibility and patient decision satisfaction.

There are limitations to the PhyReCS and our study in general. First, although the PhyReCS captures the strength of physician recommendations, it does not capture other aspects of the recommendation, such as the motivation for physicians' recommendations, which is an important factor when trying to fully understand physician recommendations. It also does not capture whether recommendations were solicited, which is known to influence how people perceive advice.[22] Second, we did not collect other measures that would help to establish concurrent validity, such as which treatment(s) physicians believed that they recommended during the appointment. Third, our study has limitations in terms of the generalizability of our results. For example, the study was conducted in the Veterans Affairs system, where patients are older, sicker, and poorer on average,[23] which may affect how physicians give recommendations. The PhyReCS may need to be adjusted with other patient populations; for example, the boundaries between recommendation scores may need to be adjusted when physicians are interacting with patients of a higher socioeconomic status. In addition, all patients (and most physicians) were male; given differences in communication styles between male and female physicians,[24] future research is needed to examine the reliability and validity of the PhyReCS in clinical settings with female patients and/or physicians. Finally, although we have evidence that our scale is valid, it is possible that another scale would have done an even better job of capturing physician recommendations. Future research is needed to optimize the scale.

In conclusion, although clinical interactions within early-stage prostate cancer are nuanced and complex, the PhyReCS makes it possible to capture how physicians recommend multiple treatment options with high reliability and validity. We feel the PhyReCS could allow researchers to more fully examine physician recommendations, an area with significant substantive and theoretical importance.

ACKNOWLEDGMENTS

We would like to thank Haley Miller, Margaret Oliver, Elizabeth Reiser, and Biqi Zhang for their help coding. We would also like to thank Valerie Kahn and Daniel Cannochie for their help in data collection and project management.

REFERENCES

1. Ranerup A. Transforming patients to consumers: evaluating national healthcare portals. Int J Public Sect Manag 2010;23(4): 331–9.

2. Thompson I, Thrasher JB, Aus G, et al. Guideline for the management of clinically localized prostate cancer: 2007 update. J Urol 2007;177(6):2106–31.

3. Barry MJ, Edgman-Levitan S. Shared decision making: the pinnacle of patient-centered care. N Engl J Med 2012;366(9):780–1.

4. Davison B, Breckon E. Factors influencing treatment decision making and information preferences of prostate cancer patients on active surveillance. Patient Educ Couns 2012;87(3):369–74.

5. Gwede CK, Pow-Sang J, Seigne J, et al. Treatment decision-making strategies and influences in patients with localized prostate carcinoma. Cancer 2005;104(7):1381–90.

6. Hall J, Boyd J, Lippert M, Theodorescu D. Why patients choose prostatectomy or brachytherapy for localized prostate cancer: results of a descriptive survey. Urology 2003;61(2):402–7.

7. Tourangeau R, Rips LJ, Rasinski K. The Psychology of Survey Response. Cambridge (UK): Cambridge University Press; 2000.

8. Loftus E, Smith K, Klinger M, Fiedler J. Memory and mismemory for health events. In: Tanur J, ed. Questions about Questions: Inquiries into the Cognitive Bases of Surveys. New York: Russell Sage Foundation; 1992. p 102–37.

9. Kunda Z. The case for motivated reasoning. Psychol Bull 1990; 108(3):480–98.

10. Veterans Affairs Office of Research and Development. In: ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2000- [cited 2016 Jun 16]. Available from: http://clinicaltrials.gov/show/NCT00432601; NLM Identifier: NCT00432601.

11. Patton MQ. Qualitative Evaluation and Research Methods. 2nd ed. Newbury Park (CA): Sage Publications; 1990.

12. Hallgren K. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol. 2012;8(1):23–34.

13. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. Commun Methods Meas. 2007;1(1):77–89.

14. Fisher B, Anderson S, Bryant J, et al. Twenty-year follow-up of a randomized trial comparing total for the treatment of invasive breast cancer. English J. 2002;347(16):1233–41.

15. Manary M, Boulding W, Staelin R, Glickman S. The patient experience and health outcomes. N Engl J Med. 2013;368(3):201–3.

16. Boulding W, Glickman SW, Manary MP, Schulman KA, Staelin R. Relationship between patient satisfaction with inpatient care and hospital readmission within 30 days. Am J Manag Care. 2011;17(1):41–8.

17. Glickman SW, Boulding W, Manary M, et al. Patient satisfaction and its relationship with clinical quality and inpatient mortality in acute myocardial infarction. Circ Cardiovasc Qual Outcomes. 2010;3(2):188–95.

18. Botti S, Lyengar SS. The psychological pleasure and pain of choosing: when people prefer choosing at the cost of subsequent outcome satisfaction. J Pers Soc Psychol. 2004;87(3):312–26.

19. Fitzsimons GJ, Lehmann DR. Reactance to recommendations: when unsolicited advice yields contrary responses. Mark Sci. 2004;23(1):82–94.

20. Goldsmith DJ, Fitch K. The normative context of advice as social support. Hum Commun Res. 1997;23(4):454–76.

21. Botti S, McGill A. When choosing is not deciding: the effect of perceived responsibility on satisfaction. J Consum Res. 2006; 33(2):211–9.

22. Dalal RS, Bonaccio S. What types of advice do decision-makers prefer? Organ Behav Hum Decis Process. 2010;112(1):11–23.

23. Oliver A. The Veterans Health Administration: an American success story? Milbank Q. 2007;85(1):5–35.

24. Roter DL, Hall JA. Physician gender and patient-centered communication: a critical review of empirical research. Annu Rev Public Health. 2004;25:497–519.