SOCIAL SCIENCES

Precision communication: Physicians' linguistic adaptation to patients' health literacy

Dean Schillinger[1,2]*†, Nicholas D. Duran[3]†, Danielle S. McNamara[4], Scott A. Crossley[5], Renu Balyan[6], Andrew J. Karter[2]

Little quantitative research has explored which clinician skills and behaviors facilitate communication. Mutual understanding is especially challenging when patients have limited health literacy (HL). Two strategies hypothesized to improve communication include matching the complexity of language to patients' HL ("universal tailoring"); or always using simple language ("universal precautions"). Through computational linguistic analysis of 237,126 email exchanges between dyads of 1094 physicians and 4331 English-speaking patients, we assessed matching (concordance/discordance) between physicians' linguistic complexity and patients' HL, and classified physicians' communication strategies. Among low HL patients, discordance was associated with poor understanding ($P = 0.046$). Physicians' "universal tailoring" strategy was associated with better understanding for all patients ($P = 0.01$), while "universal precautions" was not. There was an interaction between concordance and communication strategy ($P = 0.021$): The combination of dyadic concordance and "universal tailoring" eliminated HL-related disparities. Physicians' ability to adapt communication to match their patients' HL promotes shared understanding and equity. The 'Precision Medicine' construct should be expanded to include the domain of 'Precision Communication.'

INTRODUCTION

Physician-patient communication is a fundamental pillar of care that influences patient satisfaction, as well as health care quality, safety, and outcomes (1–3), particularly in chronic diseases such as diabetes mellitus (1, 4, 5). Recently, measurement of patients' experiences of communication has gained importance in the health policy arena, taking the form of initiatives that require public reporting of providers' performance on patient experience measures or that incentivize providers to improve interpersonal quality care (6). One of the central communicative roles of physicians is to achieve mutual understanding or "shared meaning" with patients under their care (7–9). Shared meaning, while having merit in its own right, is a determinant of a range of communication-sensitive processes and health care outcomes that emanate from a combination of physicians' eliciting and explanatory skills (9–12). Determining whether physicians' communication styles are associated with patients' understanding is of tremendous clinical and public health significance.

The average primary care physician will engage in >100,000 clinical encounters in his or her career (13). These encounters traditionally have been conducted in large part through the medium of communication, including verbal and nonverbal forms of communication. However, primary care physicians and their patients are increasingly communicating digitally via patient portals, transmitting clinical content and deliberating on health-related decisions via written email exchanges [secure messages (SMs)] (14). Research on

computer-mediated communication, however, has been lacking (3). Despite the ubiquity and importance of patient-physician communication to individual health and public health, little empirical research has been conducted—relative to the significance of the issue—to inform practice or policy.

The National Institutes of Health has invested heavily in the Precision Medicine Initiative, whose goal is to enable health care providers to tailor treatment and prevention strategies to people's unique characteristics (15). To date, the overwhelming majority of precision medicine research has focused on tailoring treatments to people's unique genetic makeup, microbiome composition, health history, and lifestyle. This has required incorporating many different types of "big data"—from genomics to metabolomics to the microbiome. Despite the centrality of shared meaning, little empirical research has been conducted to examine physician-level skills and practice styles that either inhibit or facilitate the achievement of shared meaning and patients' understanding, or explore differential effects of such physicians' behaviors across important patient subgroups (16). To our knowledge, no research has analyzed linguistic data derived from clinicians' and patients' communication exchanges as a means to advance precision medicine through precision communication.

Limited health literacy (HL) impedes the ability of physicians and patients to achieve shared understanding, imparting a barrier to patients' learning and understanding across numerous communication domains, including reading, writing, listening, speaking, and interpreting numerical values (17). More than 90 million U.S. adults are estimated to have limited HL, a problem that is more common among certain racial and ethnic groups, including non-Hispanic Blacks, Hispanics, and Asian/Pacific Islander subgroups (18). These groups are also at higher risk of acquiring type 2 diabetes and suffering from its complications. More than 30 million U.S. adults are living with diabetes, and at least one-third have limited HL skills (19, 20). Poor communication exchange appears to be in the causal pathway between limited HL and worse health outcomes (18, 21–23); limited HL is a potentially modifiable explanatory factor in the

[1]UCSF Division of General Internal Medicine and Heath Communications Research Program at the Center for Vulnerable Populations, Zuckerberg San Francisco General Hospital and Trauma Center, San Francisco, CA, USA. [2]Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA. [3]School of Social and Behavioral Sciences, Arizona State University, Glendale, AZ, USA. [4]Department of Psychology, Arizona State University, Tempe, AZ, USA. [5]Department of Applied Linguistics/ESL, College of Arts and Sciences, Georgia State University, Atlanta, GA, USA. [6]Department of Mathematics, Computer and Information Science, State University of New York, Old Westbury, NY, USA.
*Corresponding author. Email: dean.schillinger@ucsf.edu
†These authors contributed equally to this work and are co-first authors.

relationship between social determinants of health and chronic disease disparities (*18*).

Because patients with limited HL skills are more likely to report poorer quality of communication with their physician and are subject to health disparities as a result (*21*, *24*), a number of organizations and entities have advocated that training, quality improvement, and health equity efforts, including electronic physician-patient communication, focus on enhancing communication for this subgroup of patients (*25*, *26*). In parallel, there has been a growing focus on communication skill building in physician training programs (*27*). Descriptive studies have shown that clinicians' use of medical jargon and technical language is particularly problematic for patients with limited HL (*28*, *29*). HL experts and leading health care organizations (*30*) have advised that physicians use a "universal precautions" strategy (i.e., to always use simpler language) when communicating with their patients so as to reduce the so-called "literacy demand." Because of the complexities inherent to studying communication and because of lack of funding, few of these recommended practices and skills have been subject to rigorous research, undermining efforts to integrate and adopt them into practice. Specifically, no research has measured the extent of concordance between physicians' linguistic complexity and patients' HL or explored whether physicians tend to use simple language (universal precautions) versus adapt their written language to match the HL of their patients (universal tailoring). Last, no research has examined whether dyad-level concordance or physicians' tendency to use either strategy is associated with patient understanding.

These questions are of growing importance in contemporary U.S. health care, as digital forms of communication—such as SM (email) exchanges between patients and physicians via the electronic patient portal—are becoming a routine mode of care delivery in most health systems. Online portals embedded within electronic health records (EHRs) support communication between patients and providers via written SMs to bridge communication between in-person encounters (*14*). Many patients with limited HL are using SMs; thus, ensuring that this form of communication is helpful for this vulnerable group is priority (*31*). Analyzing SMs exchanged between physicians and patients enables an assessment of concordance in language complexity and an examination of whether concordance varies by patient HL.

We define linguistically discordant communications as exchanges occurring between two individuals who are fluent speakers of the same language (i.e., English) that are marked by each actors' use of different levels of linguistic complexity, sophistication, or cohesion, or frequent use of words and phrases derived from different lexicons or conceptual and sociocultural foundations. Little is known about the prevalence of linguistic discordance, the effects of linguistic discordance on patient comprehension, or whether these effects vary by patient characteristics. On the physician side, little is known about variation in the prevalence of physician-patient discordance across physicians; the extent of within-physician variation in discordance across their patients; the degree to which physicians are attuned to the linguistic capacities of their patients and tailor communication, accordingly; and the consequences of such tailoring on patient comprehension.

The current study uses data from a first-of-its-kind set of analyses generated from the ECLIPPSE Study (Employing Computational Linguistics to Improve Patient-Physician Secure email Exchange) (*11*, *32*, *33*), harnessing an unprecedented linguistic dataset that contains nearly a quarter million patient-physician SMs exchanged between >5000 dyads via an online patient portal. Leveraging advancements in machine learning and natural language processing (NLP), our interdisciplinary team developed and validated methods for automatically and accurately measuring patients' HL and physicians' language complexity from online SMs, classifying them each as high or low (*32*, *34*). One of the goals of the ECLIPPSE Study is to measure the appropriateness of physician's linguistic complexity for a particular patient's HL and determine whether concordance is associated with patient understanding. A longer-term goal is to enable real-time feedback to physicians in order that they might adapt their language to be better understood. In the current study, we carried out a retrospective analysis of a cohort of ethnically diverse, English-speaking patients with diabetes and their primary care physicians to determine the prevalence of linguistic concordance among physician-patient dyads and examined the extent to which it was associated with patients' self-reported understanding of their physicians' communication. This study had five related objectives. First, we quantified dyadic concordance and discordance by assessing the prevalence of the four possible combinations of binary characterizations of physician complexity and patient HL. Second, we assessed the relationship between dyad-level discordance and patients' reports of their understanding of their physician—measured using the physician communication subscale of the previously validated Clinician & Group Consumer Assessment of Healthcare Providers and Systems (CG-CAHPS) survey (*35*, *36*)—examining whether any association might vary by patient HL. Third, we characterized the communication strategies of physicians to study the extent to which physicians adapt their written language complexity to match the HL of their patients. Fourth, we determined whether there are associations between physicians' communication strategies and patient understanding. Last, we explored whether the combination of dyad-level concordance and strategy has synergistic effects in generating shared meaning.

RESULTS

The dataset for our initial statistical analyses included 4331 unique patients and 1143 primary care physicians. The sample was diverse with respect to race/ethnicity, sex, and educational attainment (Table 1), including patients with low (36.0%) and high HL (74.0%). Primary care physicians had between 1 and 18 patients (mean, 3.96; median, 3), with a majority (60.1%) having a mix of patients with high and low HL. Overall, physicians interacted with fewer low HL patients (mean, 1.43; SD, 1.37) than high HL patients (mean, 2.53; SD, 2.07). Across all dyads, patients sent a total of 120,387 SMs over the 10-year time frame, with each patient sending, on average, 22.08 SMs (median, 14; range, 1 to 220). The average length of the aggregated patient SMs was 1277 words (median, 784; range, 50 to 8728). Physicians sent a total of 116,739 SMs, with each physician sending, on average, 21.41 SMs (median, 14; range, 1–210). The average length of the aggregated physician SMs was 1355.99 words (median, 785; range, 70 to 8335).

Note that it is the language contained in each patient's aggregated message set that was used in a machine learning classification procedure to designate a patient's HL as high or low. Likewise, the language contained in a physician's correspondence with the patient was used to determine high or low complexity as it relates to the physician's communications (greater detail on these procedures is provided in Materials and Methods).

Table 1. Characteristics of patients in overall sample. Key patient demographics used as covariates in statistical models, stratified by HL level. The P values reported in the table are based on appropriate tests (t test, analysis of variance) of the equality between groups.

	Total (N = 4331)	HL level		P
		Low HL (n = 1560)	High HL (n = 2771)	
Age, mean ± SD)	57.2 ± 10	56.6 ± 10.2	57.6 ± 9.8	0.001
Women	1930 (44.6)	688 (44.1)	1242 (44.8)	0.671
Race/ethnicity				<0.001
White, non-Hispanic	1407 (32.5)	431 (27.6)	976 (35.2)	
Black, non-Hispanic	582 (13.4)	237 (15.2)	345 (12.5)	
Hispanic	572 (13.2)	262 (16.8)	310 (11.2)	
Asian	1348 (31.1)	484 (31)	864 (31.2)	
Other	422 (9.7)	146 (9.4)	276 (10)	
Education				<0.001
No degree	402 (9.2)	188 (12.0)	214 (7.7)	
GED/high school	995 (23)	413 (26.5)	582 (21)	
Some college or more	2934 (67.7)	959 (61.5)	1975 (71.3)	
Comorbidity score (Charlson index)	2.2 ± 1.5	2.2 ± 1.6	2.1 ± 1.5	0.067

Subhead 1: Grouping by dyadic-level SM concordance

Given that HL is associated with general linguistic and literacy skills (37–39), health communications deemed overly complex for struggling readers would be discordant with the needs of low HL patients and may undermine their understanding. Conversely, when the ability of low HL patients is met with simpler language marked, in part, by more familiar words and greater text cohesion (i.e., word and semantic overlap between sentences and messages), there would be greater SM concordance. Dyadic-level SM concordance could also be beneficial for high HL patients. Insofar as patients with higher HL show evidence of greater writing proficiency and clarity of health communication and advanced reading comprehension abilities (40, 41), they are more likely to understand a physician's more complex language and may expect more detailed explanations. Concordance in either case represents a meshing of physicians' communicative language with patients' communicative needs and abilities as determined by their HL. We considered dyadic-level SM concordance to occur when the complexity level of physicians' SM language (high or low) matched the HL level exhibited in patients' SM language (high or low). We classified each physician-patient dyad as concordant [(i) low complexity/low HL and (ii) high complexity/high HL] or discordant [(iii) high complexity/low HL and (iv) low complexity/high HL].

Across all 4331 physician/patient dyads, the prevalence of discordance (47.19%; 2044 pairs) was lower than concordance (52.81%, 2287 pairs). Stratifying by patients' HL, the proportional frequency of discordance (relative to concordance) was higher among patients with low HL compared to high HL (P < 0.001). Of the 1560 patients classified as low HL, 821 (52.62%) had a discordant physician (i.e., high complexity), whereas among the 2771 patients classified as high HL, 1223 (44.14%) had a discordant physician (i.e., low complexity).

Subhead 2: Association between dyadic-level SM concordance and patient-reported understanding

Overall, 475 (10.97%) of all patients reported "poor" understanding of their health care provider, and this lack of understanding was more prevalent among low versus high HL patients (13.65% versus 9.46%, P < 0.001). Among lower HL patients, those with a discordant physician were more likely to report poor understanding of their physician [adjusted odds ratio (OR), 1.39; P = 0.048] (Table 2). Among higher HL patients, discordance was not significantly related to patients' understanding (adjusted OR, 1.12; P = 0.418). In both the low and high HL models, non-White patients and those with lower educational attainment were also more likely to report poor understanding of physician communication.

Subhead 3: Physicians' communication strategies and patient-reported understanding

In addition to dyadic-level SM concordance, we examined whether physicians adaptively adjust their complexity to match the language of their patients and, if so, whether different physician strategies in this regard might also be associated with improved understanding. There may be a subset of physicians who tend to adopt no particular pattern of linguistic matching, wherein achieving concordance with a given patient is merely a random event. There may be other physicians who, in response to the prevalence of limited HL, tend to purposefully use low-complexity language in all interactions across all their patients, approximating the "universal precautions" strategy, wherein concordance tends to occur for only low HL patients (25, 42). In addition, there may also be physicians whose communication strategies are more adaptive and flexible, adopting a "universal tailoring" approach by tending to use lower-complexity language with their low HL patients and higher-complexity language with their high HL patients. These physicians demonstrate a greater attunement to the communicative abilities of their patients.

We classified communication strategy by categorizing physicians' distribution of language matching across their high and low HL patients. This pattern, hereafter referred to as a "tailoring signature," was then compared against ideal distributions (i.e., gold standard benchmarks) of possible communication strategies via Euclidean distance, where scores closer to 0 indicate greater resemblance to the ideal (see Fig. 1 and Methods and Materials for a more detailed explanation). While other strategies exist, we focus on universal precautions and universal tailoring given that these two approaches are often recommended for maximizing communication and represent a contrast between nonadaptive and adaptive strategies for adjusting language complexity associated with the perceived needs of low HL patients versus the perceived HL of all patients.

To establish patterns of concordance and generate unique physician tailoring signatures, we relied on a subset of 420 physicians who had at ≥2 low HL and ≥ 2 high HL patients (N = 2660). Among this

Table 2. Results of dyadic-level concordance and patient-reported understanding. Logistic regression models were specified to interpret the main fixed effect of "dyad [discordant]." The odds of having the outcome poor understanding of physician communication is either increased (odds ratio above 1) or decreased (odds ratio below 1) in discordant dyads (referent level) relative to concordant dyads. The models were adjusted for age, sex, race, and college education, with the referent level for interpreting the odds ratio placed in brackets. Standardized continuous variables indicated by (z). CI, confidence interval; ICC, intraclass correlation.

Predictor	Low HL patients			High HL patients		
	Odds ratio	95% CI	P	Odds ratio	95% CI	P
Dyad [discordant]	1.385	1.003–1.911	0.048	1.117	0.855–1.460	0.418
Covariates						
Age(z)	0.348	0.135–0.896	0.029	3.817	1.629–8.943	0.002
Sex [M]	0.966	0.703–1.326	0.829	0.991	0.759–1.294	0.945
Race [White]	0.643	0.438–0.945	0.025	0.476	0.347–0.653	<0.001
Charlson Index(z)	1.045	0.794–1.376	0.754	1.031	0.817–1.302	0.796
Some College [yes]	0.643	0.468–0.884	0.007	0.591	0.453–0.771	<0.001
Random effects						
$\tau00$ phys_ID		0.656			0.255	
Intraclass correlation (ICC)		0.166			0.072	
Marginal/conditional R^2		0.035/0.196			0.070/0.136	

sample of physicians, the prevalence of discordance (53.37%) with low HL patients was greater than concordance; among high HL patients, the prevalence of discordance (42.55%) was lower than concordance, $p < 0.001$. The most prevalent communication behavior was closest to neither of the two recommended strategies, but rather the fourth strategy in Fig. 1 that tailors only to high HL patients. In comparing the two recommended strategies, the median communication strategy resemblance score (i.e., the Euclidean distance between physicians' tailoring signatures and the gold standard benchmarks for each strategy) was similar for universal precautions and universal tailoring. However, more physicians resembled universal precautions (closer to 0) and showed greater variation across the possible range of values (0 to 2).

Using the resemblance scores for each physician, we then separately evaluated the likelihood that their patients experienced poor understanding as a function of the two recommended strategies (Table 3). We observed no significant relationship between the universal precaution resemblance scores and patient understanding for either low or high HL patients and no interaction between universal precaution strategy and patients' HL on patient-reported understanding. In contrast, we observed a significant relationship between the universal tailoring resemblance scores and patient understanding, such that patients whose physicians use strategies that deviate from universal tailoring were more likely to report poor understanding of their physician. Specifically, for each one integer unit increase in Euclidean distance, the adjusted odds of the patient experiencing poor understanding increased by a factor of 1.19, $P = 0.01$. Again, no interaction was observed between universal tailoring and HL level. In both models, non-White patients and those with lower educational attainment were more likely to report poor comprehension, independent of physicians' resemblance scores.

Subhead 4: Effects of combining dyadic-level SM concordance and physicians' communication strategy

Last, we examined the combination of the components of dyadic-level concordance and physicians' communication strategies. Starting

with models involving universal precautions, we again predicted poor patient understanding of physician, but now with an interaction term between dyadic-level concordance and physicians' resemblance score (i.e., Dyad [Discordant] * "Univ Precaution" Score). Models for both low and high HL patients did not reveal any findings of significance. Next, for models involving universal tailoring predicting poor patient understanding of physician, the interaction term between dyadic-level concordance and physicians' resemblance score (i.e., Dyad [Discordant] * "Univ Tailor" Score) was statistically significant for low HL patients only [$\chi^2(1) = 5.94$, $P = 0.01$]. Low HL patients experiencing dyadic concordance reported better understanding of their physicians' communication when those physicians more closely used universal tailoring ($b = 0.37$, SE = 0.16, $z = 2.34$, $P = 0.02$). In contrast, when experiencing dyadic discordance, having physicians who used universal tailoring appeared to generate little additional benefit ($b = -0.14$, SE = 0.14, $z = -0.97$, $P = 0.33$).

DISCUSSION

One of the most essential objectives of clinical communication is the achievement of mutual understanding between physician and patient, a process outcome that theoretically results when patients and physicians use a common language and engage in attentive and interactive verbal exchanges that, together, can enable the co-construction of shared meaning (7, 12, 43, 44). The presence of shared meaning, in turn, is a mediator for a number of key functions of the clinical encounter, such as physicians' gathering of patient symptoms and barriers to management and patients' understanding of their diagnoses and treatment plans (11). This has been best studied in chronic disease care—including in diabetes—where physician-patient miscommunication has been shown to both be common and to have untoward clinical consequences (9, 45, 46). A few small interventional studies have targeted clinician behavior to achieve shared understanding, with promising results. Approaches have included use of the interactive "teach-back method" (22) visual aids combined
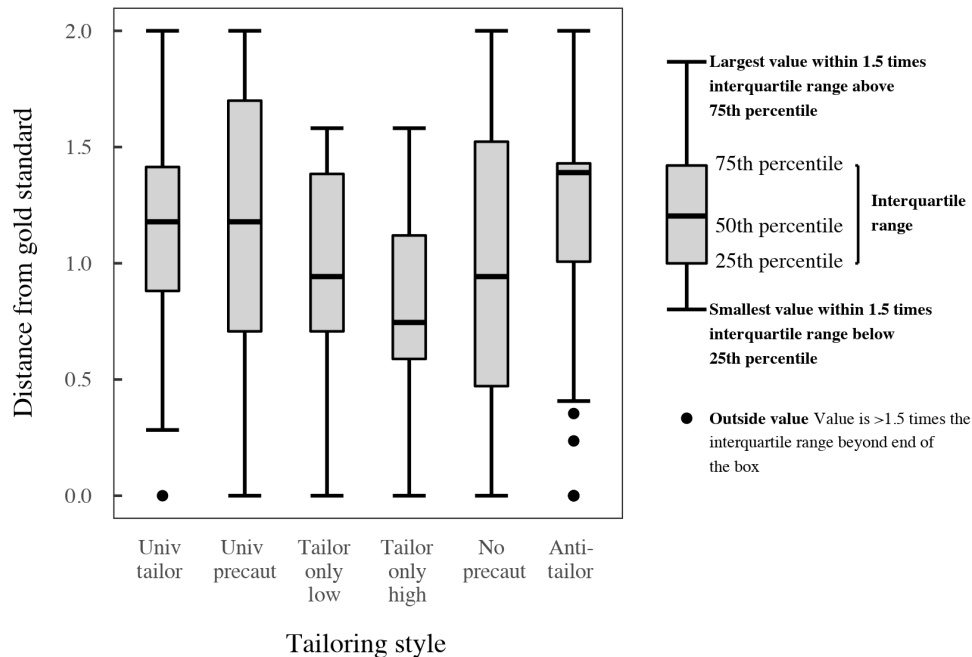
Fig. 1. Box-and-whisker plot for the distribution of physicians based on their communication strategy resemblance scores. "Distance from gold standard" (i.e., resemblance score) reflects the Euclidean distance between physicians' tailoring signatures and idealized gold standard representations of possible communication strategies (e.g., for universal tailoring, the idealized is 100% of all low-complexity messages with low HL patients and 100% of high-complexity messages with high HL patients; please see Materials and Methods for more detailed explanation of how distance scores are generated). In addition to showing distributions for universal tailoring ("Univ tailor") and universal precautions ("Univ precaut"), the figure also includes the four other mutually exclusive strategies that are conceptually possible. Resemblance to "Tailor only low" corresponds to physicians who use low-complexity language with only low HL patients but are inconsistently concordant and discordant with high HL patients. "Tailor only high" corresponds to physicians who use high-complexity language with only high HL patients but are inconsistently concordant and discordant with low HL patients. "No precaut" corresponds to physicians who consistently use high-complexity language for all patients. "Anti-tailor" corresponds to physicians who consistently use high complexity with low HL patients and use low complexity with high HL patients.

Table 3. Results of physicians' communication strategies and patient-reported understanding. (**A**) Independent effect of physicians' universal precautions resemblance score. (**B**) Independent effect of physicians' universal tailoring resemblance score. To interpret either effects, the odds of having the outcome "poor understanding" is either increased (above 1) or decreased (below 1) as communication strategy resemblance scores increase (i.e., less resemblance to the gold standard benchmarks). CI, confidence interval.

A.

Predictors	Universal precautions	
	Odds ratio	95% CI
"Univ precaution" score	0.927	0.812–1.058
Covariates		
Age(z)	1.227	0.558–2.696
Sex [M]	0.908	0.704–1.171
Race [White]	0.579	0.422–0.794
Charlson index(z)	1.141	0.921–1.415
Some college [Yes]	0.654	0.507–0.844
Random effects		
$\tau 00$ phys_ID	0.217	
ICC	0.062	
Marginal/conditional R^2	0.047/0.106	

B.

Predictors	Universal tailoring	
	Odds ratio	95% CI
'Univ tailor' score	1.191	1.040–1.365
Covariates		
Age(z)	1.177	0.537–2.578
Sex [M]	0.908	0.705–1.170
Race [White]	0.579	0.423–0.793
Charlson Index(z)	1.152	0.930–1.428
Some college [Yes]	0.655	0.509–0.844
Random effects		
$\tau 00$ phys_ID	0.182	
ICC	0.052	
Marginal/conditional R^2	0.054/0.103	

with a teach-back, and linguistic matching or tailoring (16). We attempted to address fundamental questions related to the potential importance of linguistic tailoring by applying a series of innovative approaches. First, we used advanced computational linguistics, harnessing an unprecedented, large linguistic dataset containing 10 years of written communications to analyze both patients' and physicians' SMs, classifying them as reflecting low or high HL or low or high linguistic complexity, respectively, based on recently developed and validated NLP and machine learning algorithms (32, 34, 47). On the basis of these classifications, we characterized each physician-patient dyad as either concordant or discordant. These linguistic data were complemented by patient-level data—collected via questionnaire and administrative sources—that provided relevant control variables as well as the primary outcome variable: the CG-CAHPS item. This item assesses patients' reports of the extent to which, over the last 12 months, their physicians and health care providers explained things in a way they could understand (36). This allowed us to (i) determine the prevalence of dyadic discordance, (ii) measure the independent effects of linguistic discordance on patients' reports of comprehension, and (iii) explore whether such effects vary by patient HL. Furthermore, by applying a novel technique, we were able to (iv) determine the extent of within-physician variation in communication strategies and how closely physicians' strategies approximate the ideals of "universal precautions" and "universal tailoring," (v) examine whether these two communication strategies are associated with patient reports of comprehension, and (vi) explore whether there were combined benefits to having dyad-level concordance with a physician who has a particular communication strategy. The methods used in this study represent significant advances in the field of health communication science.

First, we found that only about half of all dyadic communications were concordant and that patients with low HL experienced concordance at significantly lower rates than did high HL patients. Second, dyadic discordance was independently associated with patient reports of poor comprehension of their physicians only among low HL patients. Third, we observed within-physician variation in the extent to which physicians match their language to the HL levels of their patients, finding that physicians deploy unique communication strategies across their patients. Some have strategies that approximate the universal precautions approach, others have strategies that approximate the universal tailoring approach, and, still, others have consistently high-complexity strategies (see Fig. 1). Comparative analyses of the two recommended physician strategies revealed that, while physicians' strategies more closely resembled universal precautions, only universal tailoring was associated with better patient understanding. Fourth, we found that—for patients with low HL only—physicians' resemblance to the universal tailoring strategy moderated the effect of dyad-level discordance on communication. Specifically, for a low HL patient experiencing dyadic concordance, being cared for by a physician who also used the more adaptive strategy was associated with optimal understanding. For example, among low HL patients who experienced dyadic concordance, being cared for by a physician who had a universal tailoring resemblance score within the 0 to 25th percentile range (i.e., most likely to resemble universal tailoring) resulted in a similarly high likelihood of reporting good understanding of their physicians' communication when compared to patients who had high HL and dyadic concordance. Thus, in our sample, physicians' adoption of a strategy closely resembling universal tailoring, when combined with achieving

concordance at the level of the dyad, appears to eliminate HL-related disparities in patients' understanding.

Our findings provide confirmatory evidence that, for patients with limited HL, physicians' expressive skills—often referred to as "plain language" skills—when applied to written communications have strong potential to improve patient comprehension of physicians' clinical discourse. In any given patient with high HL, physicians' expressive skills, defined here in terms of their use of plain language versus complex language with that patient, are not significantly associated with patients' reports of understanding. However, we found that physicians' use of a universal tailoring strategy—one that requires a physician to be "attuned" to the HL of their patients and tailor and match their expressive language accordingly—is positively associated with patients' understanding of physician communication across HL levels. While we described the universal tailoring as an expressive skill, we believe that it is best thought of as an interactional communication competency since it requires an ability to gauge one's language in relation to the others' language. The process of being attuned to the HL of a patient likely reflects a higher-order function that combines the mindful and attentive reading of written communication (a form of receptive communication) with an intent and an ability to adapt one's expressive communication to match that of the patient. From this perspective, our results support the practice of mindful engagement in the clinical encounter—in this case, from a linguistic complexity standpoint—as a means to achieve shared meaning (7, 16, 48, 49).

Physicians' use of overly complex language and patients' frequent confusion or partial understanding is a problem that is so widely recognized as to almost require no proof. However, this study represents the first sizable empirical study of linguistic discordance in the medical literature, and the first observational study of its size to examine associations between dyadic discordance and physicians' communication strategies on a patient-reported outcome. Two small studies of physicians' online communications (one a simulation study and another a study of an "Ask the Doctor" website) suggested that some physicians display a tendency to adapt their language to the medical sophistication of patients' messages (50, 51). A small, randomized trial, whose objective was to examine the effects of linguistic matching when verbally communicating about potentially sensitive topics (patients' sexual or excretory function), demonstrated that concordance was associated with higher ratings of patient satisfaction, rapport, communication comfort, distress relief, and intent to adhere to treatments (16).

One notable aspect of this study is that it examined online written communication exchange. SM exchange is rapidly expanding as a means for patients and diverse members of the care team to communicate. In many health systems, SM has become a standard of care. Patients and providers are messaging at higher volumes over time. For example, at Kaiser Permanente Northern California (KPNC), >3/4 of patients are currently registered for the portal, and more than half were routinely sending SMs by 2014; the rate of SM use continues to grow. In another health care system, SMs now represent about half of the number of in-person visits and more than double the number of telephone visits (14). SM exchange is not uncommon among patients with limited HL, and rates of uptake are accelerating. In a large integrated system, between 2006 and 2015, the proportion of those with limited HL who used the portal to engage in ≥2 SM threads increased nearly 10-fold (from 6 to 57%), as compared to a fivefold increase among those with adequate HL (13 to

74%) (*31*). While additional members of the care team are increasingly being recruited as "first responders" to patients' SMs, physicians still represent the dominant segment of the workforce that engages in SM exchange (*31*).

Our measure of concordance was based on analyses of written communications. This potentially raises understandable concerns both about its relevance to clinical practice and its generalizability to verbal exchange. Nonetheless, the study sheds light on a number of previously unexplored aspects of communication that likely are relevant to verbal communication in clinical practice. At the patient level, there is reason to believe that the HL measure we used, derived from written communications, is correlated with verbal communication skills (*37*–*39*, *47*). When considering both patients' and physicians' communications, it is well established that oral language and writing are closely connected in a general way—with those who have well-developed oral language also having more well-developed writing (*52*). More particularly, writing appears to draw on verbal language, such as in the development of text cohesion, a linguistic index that was common to both the patient HL and physician linguistic complexity measures used in our study. In addition, language studies have shown parallels between verbal and computer-mediated written communication (*53*). Last, we found an association between linguistic features of written communication and patients' reports of understanding of physicians' communication, a measure that traditionally reflects physicians' face-to-face, verbal communication (*36*). On the one hand, this may have led us to underestimate the association between discordance and understanding. On the other hand, it suggests that discordance present in one medium may be associated with discordance in another.

Our outcome measure was derived from CG-CAHPS, a valid and reliable tool that measures components of patient-centeredness and engagement (*36*). While it enjoys broad application across research, clinical quality improvement, and health system performance efforts, the CG-CAHPS item provides only a proxy for shared meaning in communication exchanges. We selected this item because it is theoretically most directly aligned with the downstream consequences of linguistic discordance. A large body of literature involving millions of patients in primary care settings reveals that lower ratings of patient reports of communication using the CG-CAHPS measure are associated with a range of objectively measured communication-sensitive outcomes, including health-related behaviors (poor medication adherence), health outcomes, health care utilization, and costs (*5*, *36*), as well as communication disparities by HL level (*21*). Overall, patients in our study reported high rates of understanding of their physicians. That reports of physician communication were positively skewed is a well-known phenomenon with measures of physician satisfaction in general, including this instrument. However, prior research using the CG-CAHPS item among lower-income patients with diabetes from a safety net health system reported lower rates of understanding of physicians' communication, especially among those with low HL (*21*). While the CG-CAHPS item has been shown to overestimate patient understanding relative to direct measures of comprehension, it is unlikely to differentially do so by patient HL level (*28*). Given the limitations of the CG-CAHPS, future research should attempt to further validate our NLP-based measures of concordance through direct measurements of patient comprehension. In addition, our observational study design does not allow making causal inference between linguistic discordance and poor understanding. Carefully designed studies that intervene on physicians' communication will be needed to make judgments regarding causation.

Our study has a number of additional limitations. First, while we controlled for a number of covariates, linguistic concordance may be a marker for physician or dyadic attributes associated with patient experience. These include physician gender, experience, duration of the primary care relationship, social concordance, or a broader tendency on the part of the physician to level the hierarchy- and power-related dynamics intrinsic to the doctor-patient relationship (*54*). Second, our method of assessing linguistic content was agnostic to meaning; it is possible that concordance in meaning may have been achieved in the absence of linguistic concordance. Moreover, our assessment of physicians' linguistic complexity was based on a single measure computed over numerous patient-directed messages sent over multiple years. On the basis of theories of interpersonal interaction and shared meaning, physicians' linguistic complexity is predicted to vary in response to patients' behaviors within and across messages at more granular time scales (*44*, *55*). We are carrying out qualitative research of SMs to identify interactive content that enables or demonstrates evidence of shared meaning at the level of conversational turns. Third, our determination of concordance was based on measures of physicians' linguistic complexity and patients' HL that used NLP-based indices that tapped into similar linguistic domains. The quality of language production (in the case of patients' SMs) and the ease with which language can be read and understood by patients who may struggle with reading (in the case of physicians' SMs) have distinct theoretical underpinnings, which explains why the indices that were used to generate these measures did not have perfect overlap. As a result, dyadic concordance should not be misconstrued to simply mean that physicians' and patients' SMs had identical linguistic attributes or perfect linguistic alignment. Fourth, we found that non-White status was associated with poor comprehension even after accounting for discordance, communication strategy, and other potential confounders. This suggests that other mediators—including interactivity, trust, discrimination, and patient-centeredness (*56*)—are salient and need to be studied if equity in communication is to be realized. However, insofar as limited HL is more common among non-White patients, our findings should not be misinterpreted to mean that linguistic matching is not beneficial among such populations. It is unlikely that these particular findings were a reflection of mismeasurement; the performance characteristics and predictive validity of the automated patient HL measure have been shown to be robust and equivalent across race/ethnicity (*57*). Our inability to measure racial or ethnic concordance between physicians and patients limited our ability to explore for an interaction with linguistic concordance on communication. Last, English proficiency has been shown to be an important barrier to patient understanding and communication. However, because only a very small proportion of our sample (2.3%) reported having any difficulties speaking or understanding English, we did not have the statistical power to examine how discordance affects the experience of patients with limited English proficiency.

The central feature of nearly all clinical encounters—be they in-person, telephonic, or online—is communication exchange, and a key objective of this communication is to arrive at a shared understanding by generating shared meaning (*49*). Shared meaning is essential for collaborative problem-solving (*58*), which is at the very heart of primary care practice, and most fields in medicine. Our study not only provides impetus for promising new research in health

Table 4. Example of tailoring signature and gold standard benchmarks for generating communication strategy resemblance scores. (**A**) A hypothetical physician's tailoring signature composed of the categorization of low- and high-complexity instances with low and high HL patients. (**B**) The gold standard benchmark vectors for idealized tailoring strategies and the Euclidean distance resemblance scores for the hypothetical physician.

A.

Physician		Physician	
Low complexity		High complexity	
Low HL	High HL	Low HL	High HL
5/6 = 0.83	7/10 = 0.70	1/6 = 0.17	3/10 = 0.30

B.

	Gold standard representations				
	Physician		Physician		
	Low complexity		High complexity		
Strategy	Low HL	High HL	Low HL	High HL	Resemblance scores
Universal precautions	1	1	0	0	0.488
Universal tailoring	1	0	0	1	1.019
Tailor only low	1	0.5	0	0.5	0.371
Tailor only high	0.5	0	0.5	1	1.094
No precautions	0	0	1	1	1.307
Anti-tailor	0	1	0	1	1.452

communication but also informs and supports ongoing efforts to train physicians in communication and sheds light on how clinicians can optimize their online communication via the patient portal. Our study, which analyzed an unprecedented number of physician-patient communication exchanges using NLP and machine learning techniques, suggests that physicians' ability to adapt to patients' language is an interactional communication competency that can promote shared meaning—a proximal communication outcome in the pathway to both intermediate and distal health outcomes—and reduce HL-related communication disparities. Our work provides additional rationale for health systems to consider ways to support the efforts of clinicians and patients to seek and ultimately arrive at shared meaning as they work to navigate complex problems.

From a broader scientific perspective, our methods and findings should serve to elevate the discipline of health communication with respect to (i) its relative importance in the broader fields of clinical and public health sciences and (ii) its potential contribution to achieving both individual- and population-level health. With rare exception, communication science has not been a focus area for most major research funding entities, even in health-related circumstances and medical conditions in which communication would seem to have disciplinary relevance. For example, while the National Institutes of Health has made major investments in the Precision Medicine Initiative (15), a majority of precision medicine research funding has used "big data" related to genomics, metabolomics, or the microbiome. To our knowledge, our research is the first to consider and successfully analyze linguistic data derived from massive amounts of clinicians' and patients' communication exchanges as a means to advance precision medicine through precision communication. Our study helps make the case for expanding the construct of "Precision Medicine" to include the domain of "Precision Communication."

MATERIALS AND METHODS
Subhead 1: Study design
We carried out a retrospective analysis of a cohort of English-speaking patients with diabetes and their primary care physicians, analyzing their email exchanges over the course of a 10-year period to determine whether there is an association between physician-patient linguistic concordance and patient reports of communication with their physician.

Subhead 2: Setting and study sample
Our study involved patients with diabetes receiving ongoing care from primary care physicians at KPNC, an integrated health care system. KPNC delivers care to an insured population that is racially and ethnically diverse and is largely representative in terms of socioeconomic status, with the exception of the extremes of income (59). KPNC has a well-developed and mature patient portal, kp.org, that first enabled SM exchange in 2005. The ECLIPPSE Study derived its sample from over 20,000 ethnically diverse patients who completed a 2005–2006 survey as part of the National Institutes of Health–funded Diabetes Study of Northern California (DISTANCE). Details of DISTANCE have been reported previously (59). The DISTANCE survey collected sociodemographic and sociobehavioral measures hypothesized to influence diabetes-related health behavior, processes of care, and outcomes. Central to the current analysis is patients' assessments of their communication experience and their physicians' communication skills.

Sampling for the ECLIPPSE Study has been previously described (47, 60). Briefly, we first extracted all SMs (N = 1,050,577) exchanged from 1 January 2006 through 31 December 2015 between DISTANCE patients and all clinicians from KPNC's patient portal; the patient portal was only available in English during the study period. For the current analyses, only those SMs sent between patients and primary care physicians were included. We excluded all SMs from patients

who did not have matching DISTANCE survey data, were written in a language other than English, and were written by proxy caregivers (*61*). The total patient sample in the ECLIPPSE Study was *N* = 9530; the total physician sample was *N* = 1165. To maintain data security of this confidential information, the surveys and SMs were stored and analyzed behind KPNC's firewalls on secure servers that prevented downloading, printing, or copying.

Subhead 3: Corpus extraction
A data subset was created from the aforementioned ECLIPPSE Study by identifying all unique physician-patient communication pairs, including only those physicians linked to the patient at the time of the DISTANCE survey. We aggregated, separately for each dyad, the SMs sent by the physician to the patient and the SMs sent by the patient to the physician. SM sets were excluded if they included fewer than 50 words or were greater than 3 SDs from the mean across all message sets. These criteria ensured sufficient linguistic content to compute stable measures and minimized outlier messages where content is potentially overwhelmed by prewritten stock language copied into messages (e.g., laboratory reports, hyperlinks, and physician autostyles).

Subhead 4: Development and generation of patients' HL score
To determine high or low HL scores for each patient, we used a machine learning classification model as previously reported in our earlier work (*34*, *47*, *60*). Here, we provide a brief synopsis of the model's development, created and validated on the basis of the larger sample of patients in the ECLIPPSE Study. First, 512 unique patients were purposively sampled from the larger pool to provide roughly even representation across race/ethnicity, age, educational attainment, and self-reported HL (based on the DISTANCE survey). Two experts in HL and language processing then evaluated these patients' SM sets that contained content sent to their primary care physicians. The experts provided holistic ratings using a 6-point Likert scale related to the perceived HL of the patients (*34*). A high score corresponded to an SM thread where there was a clear mastery of written English and an advanced ability to provide logical, coherent, and well-structured health information. A low score indicated the inverse. After establishing strong interrater agreement, the ratings were averaged to provide an overall HL score for each patient. Next, a suite of NLP tools was used to extract dozens of lexical features related to word choice, discourse features, and sentence structure—all features shown to be predictors of writing quality (*62*–*65*). To determine whether an optimized set of linguistic features could distinguish low versus high HL, machine learning was used to fit a classifier model whereby the dataset was split into a training subset for model fitting and a testing subset. The final model converged on nine established linguistic features, including text cohesion, lexical sophistication, lexical diversity, and syntax (table S1), achieving high accuracy with respect to expert ratings [area under curve (AUC) of 0.87] (*47*). When this model was extended to predict the HL of the entire ECLIPPSE sample, the automated measure demonstrated consistent predictive validity (*47*): those with low HL, compared to those with high HL, reported worse understanding of their physician and exhibited worse medication adherence, worse blood sugar control, greater severity of illness, and higher emergency room utilization. We used this prediction model to automatically generate an HL score for each patient in the current dyadic dataset.

Subhead 5: Development and generation of physicians' linguistic complexity score
Physicians' communicative complexity scores were generated by a classification model adapted from one previously reported (*32*). The basic steps of model development followed the same sequence as the HL measure (*34*). New experts in reading comprehension and HL were tasked with rating 724 unique SM threads from 592 individual physicians sent to 486 unique patients. Ratings were given on a 5-point Likert scale based on the holistic question: "How easy would it be it for a struggling reader to understand this physician message?" A high complexity score indicated that a struggling reader would spend a great deal of cognitive effort and attention in translating individual and combined printed symbols into English's spoken form (i.e., the phonetics, phonology, and morphology) or in attempting to comprehend meaning in the SM, resulting in increased challenges with recognizing individual words and building meaning across sentences. A low complexity score indicated that the physician's SM would have the opposite effect. After establishing strong interrater agreement, based on the average of the experts' scores, each physician was categorized as using high- or low-complexity language. A portfolio of NLP tools was then used to generate linguistic features shown to be critical for reading comprehension (*66*–*68*). With the linguistic features as predictors and complexity as the dependent variable, a classifier model was built using a cross-validation technique that allows training and testing on independent data. The final model, named the physician complexity profile, contained 24 linguistic features that correctly classified the SMs within the test with a good degree of accuracy (AUC, 0.75) (*32*). In applying this model to the current dyadic dataset, we were able to streamline the number of linguistic features to comprise 14 indices rather than the original 24 (table S1), including text cohesion, lexical sophistication, lexical diversity, and syntax, while retaining equivalent accuracy.

Subhead 6: Classifying physician-level communication strategies
To generate communication strategy resemblance scores, we first reduced the dataset to include only those physicians who had at least four patients, with at least two patients having low HL and two having high HL (the minimum number necessary to establish any systematic patterns in complexity use across patients of varying HL). Because the total number of patients that each physician had varied, the next step was to compute the proportion of patients with whom a physician was concordant or discordant—separately for the total number of low and high HL patients. For example, a hypothetical physician might have a total of 6 low HL and 10 high HL patients. This physician used low-complexity language with 5 of the 6 lower HL patients (5/6 = 0.83; concordant) and with 7 of the 10 high HL patients (7/10 = 0.70; discordant) (Table 4A). The physician used high complexity with one low HL patient (1/6 = 0.17; discordant) and high complexity with high HL patients (3/10 = 0.30; concordant). Together, these proportion scores create a four-element distribution vector (i.e., tailoring signature): [0.83, 0.70, 0.17, 0.30].

In the last step, the tailoring signature was compared, via Euclidean distance, to gold standard vector benchmarks. Each benchmark was composed of ideal proportion rates of discordance and concordance for each communication strategy. For example, the four-element "universal precautions" vector of [1, 1, 0, 0] corresponds to a situation where there is consistent concordance with lower HL patients

and consistent discordance with high HL patients, with no instances of high-complexity language use for any patient (Table 4B). The "universal tailoring" vector [1, 0, 0, 1] corresponds to consistent concordance across all low and high HL patients, with no instances of discordance. It would be uncommon for physicians' tailoring signature to map perfectly onto any benchmark vector; accordingly, we resisted the temptation of placing physicians in any absolute category. Instead, physicians were considered to be on a graded continuum, where the patterning of complexity matching across patients was measured continuously as being closer to one gold standard benchmark over another. The "resemblance scores" shows the Euclidean distance results of our example physician's tailoring signature with each benchmark vector. Scores closer to zero indicate greater resemblance to a particular communication strategy, whereas higher scores, to a maximum of 2.00, indicate less resemblance. Our hypothetical physician has a closer resemblance to a "universal precautions" strategy (Euclidean distance score of 0.488) than to a universal tailoring strategy (score of 1.019). However, across all possible strategies, based on the Euclidean distance from gold standards, this physician most resembles the strategy that tailors only to low HL patients (score of 0.371), i.e., consistently using low complexity with all low HL patients ([1]), never using high complexity with low HL patients ([0]), and inconsistent use of complexity with high HL patients ([0.5]).

Subhead 7: Outcome
Our outcome was derived from the physician communication subscale of the previously validated CG-CAHPS survey (*35*, *36*), obtained from patients as part of the DISTANCE survey, which reflected their experiences with the KPNC physician to whom they were assigned. We selected the item that measures patients' understanding of their physician as it represents the most salient dimension of physician-patient communication based on the objective of this study. The question asks patients to report, "in the last 12 months, how often [their] doctors or health care providers explained things in a way [they] could understand." Responses were given on an ordinal scale of 1 to 4 (never, sometimes, usually, or always). For purposes of this analysis, they were dichotomized into a quality of communication score of "poor" (combining "never" and "sometimes") versus "good" (combining "usually" and "always") (*69*). This item and its subscale have been used in related studies with diabetes patients, including as the outcome in a study demonstrating that limited HL is associated with poor physician communication (*21*) and as the predictor in a study demonstrating that poor physician communication is associated with lower rates of cardiometabolic medication and antidepressant medication adherence (*5*, *70*).

Subhead 8: Control variables
On the basis of the literature on patient-level factors associated with the quality of physician-patient communication, we included age, sex, race/ethnicity, educational attainment, and comorbidity. These were collected via the DISTANCE survey, with the exception of the Charlson comorbidity index, which was derived from EHR data (*71*). We used binary dummy variables for sex, either male or female; for race, either White or non-White; and for level of education, either high school/no degree or some form of postsecondary/college education. Patients' age and comorbidity index were included as continuous, *z*-scored variables.

Subhead 9: Statistical analysis
Because of the unique theoretical assumptions for the associations between dyadic-level SM concordance and physician-level tailoring with the outcome, we evaluated each model separately before exploring any potential synergy between the two predictors. All analyses use generalized linear mixed-effects models with a binomial outcome, predicting the likelihood of "poor" communication (coded as "1") relative to "good" (coded as "0"). Models were adjusted for the patient characteristics described above and included a random effect term for the clustered variance of patients within physician (coded as "phys_ID"). We report the odds ratios for factors in each model alongside P values derived from the z-distribution applied to the Wald t values for each factor, with P values <0.05 considered statistically significant. The t-distribution begins to approximate the z-distribution as the number of observations increases (all our models include at least 1000 data points). For analyses involving an interaction, we computed a model with and without the interaction and compared model fits using a likelihood ratio test. Where the interaction is not statistically significant, we report results from the main effects model. Approximation of overall effect size is reported as variance explained by fixed effects alone, marginal R^2, and with the inclusion of random factors, conditional R^2 (*72*). All analyses were performed with R statistical software using the lme4 mixed effects package (version 1.1-23).

For our analyses involving dyadic-level SM concordance, the fixed effect was at the level of physician-patient dyad, dummy coded with the referent group set to "concordant" as 0 and "discordant" as 1. We performed stratified analyses a priori, separately evaluating low and high HL patients given that what constitutes concordance differs between the two groups (i.e., physician using high complexity is concordant for high HL but discordant for low HL).

For our analyses involving physician-level communication style, the fixed effects were physicians' resemblance scores for universal precautions and for universal tailoring (each z scored and mean centered), using separate models to evaluate each. For both models, we included patients' HL level, dummy coded with referent group set to "high" as 0 and "low" as 1. By including HL level, our interest was also to explore for an interaction with resemblance scores to determine whether particular communication strategies have a unique impact on patients' communication experience for low versus high HL patients.

Last, we explored whether any possible advantage of dyadic-level SM concordance is more pronounced when the concordance occurs with a physician whose communication strategy is more aligned with the universal precautions or universal tailoring approach. In other words, we examined the moderating effect (i.e., statistical interaction) of physicians' resemblance scores on dyadic message–level concordance. Of particular interest was whether low HL patients benefit most when their physicians use lower-complexity language in combination with either a more fixed and safety-oriented communication strategy ("universal precautions") or a more flexible and adaptive communication strategy ("universal tailoring").

REFERENCES AND NOTES
1. C. Doyle, L. Lennox, D. Bell, A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open* **3**, e001570 (2013).

2. D. Riedl, G. Schüßler, The influence of doctor-patient communication on health outcomes: A systematic review. *Z. Psychosom. Med. Psychother.* **63**, 131–150 (2017).

3. S. A. Lee, R. J. Zuercher, A current review of doctor–patient computer-mediated communication. *J. Commun. Healthc.* **10**, 22–30 (2017).

4. M. A. Stewart, Effective physician-patient communication and health outcomes: A review. *CMAJ* **152**, 1423–1433 (1995).

5. N. Ratanawongsa, A. J. Karter, M. M. Parker, C. R. Lyles, M. Heisler, H. H. Moffet, N. Adler, E. M. Warton, D. Schillinger, Communication and medication refill adherence the diabetes study of Northern California. *JAMA Intern. Med.* **173**, 210–218 (2013).

6. C. M. Setodji, Q. Burkhart, R. D. Hays, D. D. Quigley, S. A. Skootsky, M. N. Elliott, Differences in consumer assessment of healthcare providers and systems clinician and group survey scores by recency of the last visit: Implications for comparability of periodic and continuous sampling. *Med. Care* **57**, E80–E86 (2019).

7. R. McCabe, P. G. T. Healey, Miscommunication in doctor–patient communication. *Top. Cogn. Sci.* **10**, 409–424 (2018).

8. R. M. Epstein, R. L. Street, The values and value of patient-centered care. *Ann. Fam. Med.* **9**, 100–103 (2011).

9. U. Sarkar, D. Schillinger, K. Bibbins-Domingo, A. Nápoles, L. Karliner, E. J. Pérez-Stable, Patient-physicians' information exchange in outpatient cardiac care: Time for a heart to heart? *Patient Educ. Couns.* **85**, 173–179 (2011).

10. E. Sagi, D. Diermeier, Language use and coalition formation in multiparty negotiations. *Cognit. Sci.* **41**, 259–271 (2017).

11. D. Schillinger, D. Mcnamara, S. Crossley, C. Lyles, H. H. Moffet, U. Sarkar, N. Duran, J. Allen, J. Liu, D. Oryn, N. Ratanawongsa, A. J. Karter, The next frontier in communication and the ECLIPPSE study: Bridging the linguistic divide in secure messaging. *J. Diabetes Res.* **2017**, 1–9 (2017).

12. R. M. Epstein, R. L. Street, Shared mind: Communication, decision making, and autonomy in serious illness. *Ann. Fam. Med.* **9**, 454–461 (2011).

13. B. White, D. Twiddy, *The State of Family Medicine: 2017* (2017);www.aafp.org/fpm.

14. J. Yakushi, M. Wintner, N. Yau, L. Borgo, E. Solorzano, Utilization of secure messaging to primary care departments. *Perm. J.* **24**, 19.177 (2020).

15. F. S. Collins, H. Varmus, A new initiative on precision medicine. *N. Engl. J. Med.* **372**, 793–795 (2015).

16. N. Williams, J. Ogden, The impact of matching the patient's vocabulary: A randomized control trial. *Fam. Pract.* **21**, 630–635 (2004).

17. N. D. Berkman, T. C. Davis, L. McCormack, Health literacy: What is it? *J. Health Commun.* **15**, 9–19 (2010).

18. D. Schillinger, Social Determinants, Health Literacy, and Disparities: Intersections and Controversies. *Health Lit. Res. Pract.* **5**, e234–e243 (2021).

19. D. Schillinger, K. Grumbach, J. Piette, F. Wang, D. Osmond, C. Daher, J. Palacios, G. D. Sullivan, A. B. Bindman, Association of health literacy with diabetes outcomes. *JAMA* **288**, 475–482 (2002).

20. S. C. Bailey, A. G. Brega, T. M. Crutchfield, T. Elasy, H. Herr, K. Kaphingst, A. J. Karter, S. Moreland-Russell, C. Y. Osborn, M. Pignone, R. Rothman, D. Schillinger, Update on health literacy and diabetes. *Diabetes Educ.* **40**, 581–604 (2014).

21. D. Schillinger, A. Bindman, F. Wang, A. Stewart, J. Piette, Functional health literacy and the quality of physician-patient communication among diabetes patients. *Patient Educ. Couns.* **52**, 315–323 (2004).

22. D. Schillinger, J. Piette, K. Grumbach, F. Wang, C. Wilson, C. Daher, K. Leong-Grotz, C. Castro, A. B. Bindman, Closing the loop: Physician communication with diabetic patients who have low health literacy. *Arch. Intern. Med.* **163**, 83–90 (2003).

23. C. Brach, D. Keller, L. Hernandez, C. Baur, R. Parker, B. Dreyer, P. Schyve, A. J. Lemerise, D. Schillinger, Ten attributes of health literate health care organizations, in *NAM Perspect.* **02** (2012);https://nam.edu/perspectives-2012-ten-attributes-of-health-literate-health-care-organizations/.

24. U. Sarkar, J. D. Piette, R. Gonzales, D. Lessler, L. D. Chew, B. Reilly, J. Johnson, M. Brunt, J. Huang, M. Regenstein, D. Schillinger, Preferences for self-management support: Findings from a survey of diabetes patients in safety-net health systems. *Patient Educ. Couns.* **70**, 102–110 (2008).

25. M. Jablow, "Say that again? Teaching physicians about plain language," *AAMC News*, 11 October 2016, pp. 1–9.

26. W. W. Lee, L. S. Sulmasy; American College of Physicians Ethics, Professionalism and Human Rights Committee, American College of Physicians ethical guidance for electronic patient-physician communication: Aligning expectations. *J. Gen. Intern. Med.* **35**, 2715–2720 (2020).

27. Strategy 6G: Training to advance physicians' communication skills (Agency for Healthcare Research and Quality); www.ahrq.gov/cahps/quality-improvement/improvement-guide/6-strategies-for-improving/communication/strategy6gtraining.html.

28. C. M. Castro, C. Wilson, F. Wang, D. Schillinger, Babel babble: Physicians' use of unclarified medical jargon with patients. *Am. J. Health Behav.* **31**, S85–S95 (2007).

29. G. Joseph, R. J. Pasick, D. Schillinger, J. Luce, C. Guerra, J. K. Y. Cheng, Information mismatch: Cancer risk counseling with diverse underserved patients. *J. Genet. Couns.* **26**, 1090–1104 (2017).

30. A. G. Brega, M. A. G. Freedman, W. G. LeBlanc, J. Barnard, N. M. Mabachi, M. Cifuentes, K. Albright, B. D. Weiss, C. Brach, D. R. West, Using the health literacy universal precautions toolkit to improve the quality of patient materials. *J. Health Commun.* **20**, 69–76 (2015).

31. A. G. Cemballi, A. Karter, D. Schillinger, J. Y. Liu, D. S. McNamara, W. Brown, S. Crossley, W. Semere, M. Reed, J. Allen, C. R. Lyles, Descriptive examination of secure messaging in a longitudinal cohort of diabetes patients in the ECLIPPSE study. *J. Am. Med. Inf. Assoc.* **28**, 1252–1258 (2020).

32. S. A. Crossley, R. Balyan, J. Liu, A. J. Karter, D. McNamara, D. Schillinger, Predicting the readability of physicians' secure messages to improve health communication using novel linguistic features: Findings from the ECLIPPSE study. *J. Commun. Healthc.* **13**, 344–356 (2020).

33. R. Balyan, S. A. Crossley, W. Brown III, A. J. Karter, D. S. McNamara, J. Y. Liu, C. R. Lyles, D. Schillinger, Using natural language processing and machine learning to classify health literacy from secure messages: The ECLIPPSE study. *PLOS ONE* **14**, e0212488 (2019).

34. S. A. Crossley, R. Balyan, J. Liu, A. J. Karter, D. McNamara, D. Schillinger, Developing and testing automatic models of patient communicative health literacy using linguistic features: Findings from the ECLIPPSE study. *Health Commun.* **36**, 1018–1028 (2020).

35. *CAHPS 2.0 Questionnaires AHCPR Pub. No. 97-R079* (Washington DC, 1998).

36. J. M. Holt, Patient experience in primary care: A systematic review of CG-CAHPS surveys. *J. Patient Exp.* **6**, 93–102 (2019).

37. S. S. Nouri, R. E. Rudd, Health literacy in the "oral exchange": An important element of patient-provider communication. *Patient Educ. Couns.* **98**, 565–571 (2015).

38. M. Schonlau, L. Martin, A. Haas, K. P. Derose, R. Rudd, in *Journal of Health Communication* (Taylor & Francis Group, 2011), vol. 16, pp. 1046–1054.

39. S. Koch-Weser, R. E. Rudd, W. Dejong, Quantifying word use to study health literacy in doctor-patient communication. *J. Health Commun.* **15**, 590–602 (2010).

40. L. K. Allen, E. L. Snow, S. A. Crossley, G. T. Jackson, D. S. McNamara, Reading comprehension components and their relation to writing. *Annee Psychol.* **114**, 663–691 (2014).

41. R. Schoonen, Are reading and writing building on the same skills? The relationship between reading and writing in L1 and EFL. *Read. Writ.* **32**, 511–535 (2019).

42. Health Literacy Universal Precautions Toolkit, 2nd Edition (Agency for Healthcare Research and Quality); https://www.ahrq.gov/health-literacy/improve/precautions/toolkit.html.

43. W. S. Horton, R. J. Gerrig, Conversation common ground and memory processes in language production. *Discourse Process.* **40**, 57–82 (2005).

44. B. M. Watson, L. Jones, D. G. Hewett, in *Communication Accommodation Theory: Negotiating Personal Relationships and Social Identities across Contexts*, H. Giles, Ed. (Cambridge Univ. Press, Cambridge, 2016), pp. 152–168.

45. D. Schillinger, F. Wang, M. Rodriguez, A. Bindman, E. L. Machtinger, The importance of establishing regimen concordance in preventing medication errors in anticoagulant care. *J. Health Commun.* **11**, 555–567 (2006).

46. J. L. Barton, J. Imboden, J. Graf, D. Glidden, E. H. Yelin, D. Schillinger, Patient-physician discordance in assessments of global disease severity in rheumatoid arthritis. *Arthritis Care Res. (Hoboken).* **62**, 857–864 (2010).

47. D. Schillinger, R. Balyan, S. A. Crossley, D. S. McNamara, J. Y. Liu, A. J. Karter, Employing computational linguistics techniques to identify limited patient health literacy: Findings from the ECLIPPSE study. *Health Serv. Res.* **56**, 132–144 (2020).

48. D. M. Zulman, M. C. Haverfield, J. G. Shaw, C. G. Brown-Johnson, R. Schwartz, A. A. Tierney, D. L. Zionts, N. Safaeinili, M. Fischer, S. Thadaney Israni, S. M. Asch, A. Verghese, Practices to foster physician presence and connection with patients in the clinical encounter. *JAMA* **323**, 70–81 (2020).

49. H. H. Clark, *Using Language* (Cambridge Univ. Press, Cambridge, 1996).

50. R. Jucks, R. Bromme, Choice of words in doctor-patient communication: An analysis of health-related internet sites. *Health Commun.* **21**, 267–277 (2007).

51. R. Bromme, R. Jucks, T. Wagner, How to refer to "Diabetes"? Language in online health advice. *Appl. Cogn. Psychol.* **19**, 569–586 (2005).

52. T. Shanahan, in *Handbook of Writing Research* (Guildford Press, 2006), pp. 171–182.

53. J. B. Walther, T. Loh, L. Granka, Let me count the ways: The interchange of verbal and nonverbal cues in computer-mediated and face-to-face affinity. *J. Lang. Soc. Psychol.* **24**, 36–65 (2005).

54. R. L. J. Thornton, N. R. Powe, D. Roter, L. A. Cooper, Patient-physician social concordance, medical visit communication and patients' perceptions of health care quality. *Patient Educ. Couns.* **85**, e201–e208 (2011).

55. D. Angus, B. Watson, A. Smith, C. Gallois, J. Wiles, Visualising conversation structure across time: Insights into effective doctor-patient consultations. *PLOS ONE* **7**, e38014 (2012).

56. D. L. Roter, S. Larson, D. Z. Sands, D. E. Ford, T. Houston, Can E-mail messages between patients and physicians be patient-centered? *Health Commun.* **23**, 80–86 (2008).

57. D. Schillinger, R. Balyan, S. Crossley, D. McNamara, A. Karter, Validity of a computational linguistics-derived automated health literacy measure across race/ethnicity: Findings from the ECLIPPSE project. *J Healthc. Poor Underserved* **32**, 347–365 (2021).

58. J. Roschellel, S. D. Teasley, in *Computer-Supported Collaborative Learning*, C. E. O'Malley, Ed. (Springer-Verlag, Berlin, 1995), pp. 69–97.

59. H. H. Moffet, N. Adler, D. Schillinger, A. T. Ahmed, B. Laraia, J. V. Selby, R. Neugebauer, J. Y. Liu, M. M. Parker, M. Warton, A. J. Karter, Cohort Profile: The Diabetes Study of Northern California (DISTANCE)—Objectives and design of a survey follow-up study of social health disparities in a managed care population. *Int. J. Epidemiol.* **38**, 38–47 (2009).

60. W. Brown III, R. Balyan, A. J. Karter, S. Crossley, W. Semere, N. D. Duran, C. Lyles, J. Liu, H. H. Moffet, R. Daniels, D. S. Mcnamara, D. Schillinger, Challenges and solutions to employing natural language processing and machine learning to measure patients' health literacy and physician writing complexity: The ECLIPPSE study. *J. Biomed. Inform.* **113**, 103658 (2021).

61. W. Semere, S. Crossley, A. J. Karter, C. R. Lyles, W. Brown III, M. Reed, D. S. McNamara, J. Y. Liu, D. Schillinger, Secure messaging with physicians by proxies for patients with diabetes: Findings from the ECLIPPSE study. *J. Gen. Intern. Med.* **34**, 2490–2496 (2019).

62. S. A. Crossley, D. S. McNamara, Say more and be more coherent: How text elaboration and cohesion can increase writing quality. *J. Writ. Res.* **7**, 351–370 (2016).

63. D. S. McNamara, S. A. Crossley, R. D. Roscoe, L. K. Allen, J. Dai, A hierarchical classification approach to automated essay scoring. *Assess. Writ.* **23**, 35–59 (2015).

64. K. Kyle, S. A. Crossley, Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *Mod. Lang. J.* **102**, 333–349 (2018).

65. S. A. Crossley, L. K. Allen, E. L. Snow, D. S. Mcnamara, Incorporating learning characteristics into automatic essay scoring models: What individual differences and linguistic features tell us about writing quality. *J. Educ. Data Min.* **8**, 1–19 (2016).

66. K. Kyle, S. A. Crossley, Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Q.* **49**, 757–786 (2015).

67. S. A. Crossley, K. Kyle, D. S. McNamara, Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behav. Res. Methods* **49**, 803–821 (2017).

68. A. C. Graesser, D. S. McNamara, M. M. Louwerse, Z. Cai, Coh-Metrix: Analysis of text on cohesion and language. *Behav. Res. Methods* **36**, 193–202 (2004).

69. B. D. Stucky, R. D. Hays, M. O. Edelen, J. Gurvey, J. A. Brown, Possibilities for shortening the CAHPS Clinician and Group survey. *Med. Care* **54**, 32–37 (2016).

70. A. M. Bauer, M. M. Parker, D. Schillinger, W. Katon, N. Adler, A. S. Adams, H. H. Moffet, A. J. Karter, Associations between antidepressant adherence and shared decision-making, patient-provider trust, and communication among adults with diabetes: Diabetes Study of Northern California (DISTANCE). *J. Gen. Intern. Med.* **29**, 1139–1147 (2014).

71. M. Charlson, T. P. Szatrowski, J. Peterson, J. Gold, Validation of a combined comorbidity index. *J. Clin. Epidemiol.* **47**, 1245–1251 (1994).

72. P. C. D. Johnson, Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods Ecol. Evol.* **5**, 944–946 (2014).