Secure Messaging with Physicians by Proxies for Patients with Diabetes: Findings from the ECLIPPSE Study

Wagahta Semere, MD, MHS[1,2], Scott Crossley, PhD[3], Andrew J. Karter, PhD[1,2,4], Courtney R. Lyles, PhD[1,2,4], William Brown III, PhD, DrPH, MA[5], Mary Reed, PhD[4], Danielle S. McNamara, PhD[6], Jennifer Y. Liu, MPH[4], and Dean Schillinger, MD[1,2,4,7]

[1]Division of General Internal Medicine, University of California, San Francisco, San Francisco, CA, USA; [2]Zuckerberg San Francisco General Hospital, San Francisco, CA, USA; [3]Department of Applied Linguistics and English as a Second Language, Georgia State University, Atlanta, GA, USA; [4]Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA; [5]Center for AIDS Prevention Studies, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA; [6]Department of Psychology, Arizona State University, Tempe, AZ, USA; [7]Center for Vulnerable Populations, Zuckerberg San Francisco General Hospital, University of California, San Francisco, San Francisco, CA, USA.

BACKGROUND: Little is known about patients who have caregiver proxies communicate with healthcare providers via portal secure messaging (SM). Since proxy portal use is often informal (e.g., sharing patient accounts), novel methods are needed to estimate the prevalence of proxy-authored SMs.

OBJECTIVE: (1) Develop an algorithm to identify proxy-authored SMs, (2) apply this algorithm to estimate predicted proxy SM (PPSM) prevalence among patients with diabetes, and (3) explore patient characteristics associated with having PPSMs.

DESIGN: Retrospective cohort study.

PARTICIPANTS: We examined 9856 patients from Diabetes Study of Northern California (DISTANCE) who sent ≥ 1 English-language SM to their primary care physician between July 1, 2006, and Dec. 31, 2015.

MAIN MEASURES: Using computational linguistics, we developed ProxyID, an algorithm that identifies phrases frequently found in registered proxy SMs. ProxyID was validated against blinded expert categorization of proxy status among an SM sample, then applied to identify PPSM prevalence across patients. We examined patients' sociodemographic and clinical characteristics according to PPSM penetrance, "none" (0%), "low" (≥ 0–50%), and "high" (≥ 50–100%).

KEY RESULTS: Only 2.3% of patients had ≥ 1 registered proxy-authored SM. ProxyID demonstrated moderate agreement with expert classification ($K = 0.58$); 45.7% of patients had PPSMs (40.2% low and 5.5% high). Patients with high percent PPSMs were older than those with low percent and no PPSMs (66.5 vs 57.4 vs 56.2 years, $p < 0.001$) had higher rates of limited English proficiency (16.1% vs 3.2% vs 3.5%, $p < 0.05$), lower self-reported health literacy (3.83 vs 4.43 vs 4.44, $p < 0.001$), and more comorbidities (Charlson index 3.78 vs 2.35 vs 2.18, $p < 0.001$).

CONCLUSIONS: Among patients with diabetes, informal proxy SM use is more common than registered use and prevalent among socially and medically vulnerable patients. Future research should explore whether proxy portal use improves patient and/or caregiver outcomes and consider policies that integrate caregivers in portal communication.

KEY WORDS: caregiving; diabetes; health information technology (health IT); health communication.

J Gen Intern Med 34(11):2490–6
DOI: 10.1007/s11606-019-05259-1

Prior Presentations: *Poster Presentation at SGIM Annual Meeting, April 11, 2018, Denver, CO*

Received November 26, 2018
Revised May 14, 2019
Accepted July 18, 2019
Published online August 19, 2019

INTRODUCTION

Currently, 34 million American adults provide care for an aging family member, with more than half of their care recipients having multiple complex care needs.[1] The National Academies recently convened experts to better understand attributes and needs of family caregivers.[2] Family caregivers experience poor physical health, emotional instability, and economic difficulties, in part, due to their caregiving burden.[3, 4] These findings and the known benefits that caregiving yields patients support caregiver-centered approaches in healthcare.[5]

Few studies have examined electronic communication between caregivers and clinicians.[5, 6] Caregivers and their care recipients express interest in using health IT to facilitate managing information and communicating with healthcare teams.[6–8] Patient portals—systems that provide patients remote access to health information and secure messaging (SM)—allow patients' proxies to communicate with providers on their behalf. This can occur formally, wherein patients assign a registered proxy (e.g., spouse, adult child, other relative, or friend) health record and SM access.[9] Informally, proxies may log in as the patient and send SMs.[10, 11] Little is known about prevalence of informal proxy use, or characteristics of patients who rely on them.

Engaging caregivers through health IT can lead to improved patient outcomes, communication with providers, and confidence in self-management.[12, 13] In diabetes, patients have frequent healthcare encounters and significant self-management and communication needs, making portal access beneficial. Yet, information is lacking regarding the mechanisms through which portals may influence patient outcomes.[14] Despite growing Internet access, older adults and those with communication barriers face challenges using patient portals that proxies might mitigate.[15–19] Describing the prevalence of proxy SM use and determining which patients rely on proxies is important for understanding how to leverage proxies and health IT to improve patient and caregiver outcomes.

The objectives of this study are (1) to develop and validate an algorithm that predicts proxy SMs by informal proxies (PPSMs), (2) to apply this algorithm to estimate the prevalence of PPSMs among a large, diverse sample of patients with diabetes, and (3) to determine which sociodemographic and clinical characteristics of patients are associated with PPSMs.

METHODS

Study Sample and Setting

This study arose from the ECLIPPSE (Employing Computational Linguistics to Improve Patient-Provider Secure Emails) project, drawing from the well-characterized Diabetes Study of Northern California (DISTANCE) sample, designed to examine social behavioral factors associated with disparities in diabetes-related care and outcomes.[20] DISTANCE was fielded from 2005 to 2006—using a combination of phone, computer, and paper survey methods—with an ethnically stratified random sample of patients in a large, integrated healthcare delivery system, Kaiser Permanente of Northern California. Among > 20,000 respondents, half were > 60 years, with diverse racial/ethnic representation (16.9% African American, 18.4% Latino, 11.4% Asian).[21] Sixty percent had some limitation in self-reported HL.[22]

Kaiser launched a patient portal in 1999, which by 2005 allowed patients to send SMs to their providers. For this study, we examined 9856 DISTANCE patients who composed ≥ 1 English-language SM to their primary care physician between July 1, 2006, and December 31, 2015.

Development and Validation of the ProxyID Algorithm

In 2006, Kaiser's portal enabled "Act for a Family Member," which allowed patients to designate a family member or other caregiver access to their personal health record. Registered proxy users could access lab results, schedule appointments, refill prescriptions, and send/receive SMs to/from providers. Outside of "Act for a Family Member," it is not known how often proxy users informally perform tasks on behalf of patients.

To identify patients using proxies to communicate with providers, we developed ProxyID, an algorithm using corpus linguistics strategies to detect words and phrases that were more likely to appear in proxy SMs compared to patient-authored SMs. We first identified registered proxy-authored SMs then randomly sampled an equivalent number of presumed patient-authored SMs. We used Wordsmith Tools 6 to identify key n-grams (i.e., words and contiguous phrases) more likely than chance to occur in registered proxy SMs compared to presumed patient-authored SMs.[23, 24] We considered n-grams from one to two words in length to be key if they occurred in registered proxy SMs significantly more or less frequently than in presumed patient-authored SMs. To ensure that the key distinction of a particular n-gram was representative of use across registered proxy SMs, we set the minimum threshold for inclusion at a range of 10% (i.e., selected n-grams had to occur in ≥ 10% of registered proxy SMs). Once we identified key n-grams, for each SM, we calculated a normed frequency for how often these n-grams appeared in the registered proxy SMs and the presumed patient-authored SMs using the Natural Language Processing (NLP) tool SiNLP.[23]

The percentage of n-grams was fed into ProxyID, which relied on machine learning to select proxy messages based on this data and patterns of n-grams in the messages. This enabled us to classify each SM as a "PPSM" or "non-proxy SM." To assess validity of these classifications, three expert assessors blinded to the predicted proxy status of the SM each read SMs from a purposive sample of 200 unique patients (100 SMs designated by ProxyID as PPSMs and 100 designated as non-proxy SMs) and, based on SM content, categorized these SMs as PPSMs or non-proxy SMs. Kappa scores were calculated to determine concordance between ProxyID and expert assessors.

Distribution of Proxy SMs

We defined three groups of SMs—one group authored by a registered proxy, a second predicted as authored by an informal proxy (PPSMs), and a third predicted as authored by the patient (non-proxy). We then examined total number of PPSMs for each patient, categorizing patients according to the distribution of their SMs estimated to be PPSMs (i.e., percent of PPSMs, in deciles, e.g. 10%, 20% …).

Patient Sociodemographic and Clinical Characteristics

We hypothesized that having barriers in communication (low HL, LEP, dementia) would be associated with reliance on proxies. The DISTANCE sample measured sociodemographic characteristics, including rates of LEP and level of HL. For LEP, patients were asked, "How often do you have difficulty understanding or speaking English?" (DISTANCE Q26); responses were dichotomized as LEP ("Always," "Often," and "Sometimes") or English proficient ("Rarely" and "Never").[25]

HL scores were determined using the mean of three validated self-reported HL measures (difficulty understanding written health information, confidence filling out medical forms independently, and need for help reading health information).[26]

We hypothesized that patients with complex diabetes care (e.g., high HA1c, insulin use, and more comorbidities) would rely on proxies. The DISTANCE survey asked about duration of diabetes diagnosis, diabetes knowledge, and self-reported episodes of hypoglycemia. Additional patient characteristics were collected from the EHR the year prior to patients sending their first SM. These included Hemoglobin A1c, insulin use, total number of medications, dementia diagnosis, Charlson comorbidity score, and measures of healthcare utilization (outpatient, inpatient, and emergency room visits).

Statistical Analysis. The n-gram frequency analyses for the ProxyID algorithm identified key words and phrases that differentiated registered proxy SMs and PPSMs. These key word and phrase indices were entered into a MANOVA analysis to identify n-grams with differences ($p < 0.05$) between registered proxy-authored SMs and PPSMs. We then conducted a stepwise discriminant function analysis (DFA) to examine how well the key words and phrases could discriminate between patients who had registered proxy SMs and those without registered proxy SMs. Based on this analysis, we generated ProxyID and examined sensitivity, specificity, negative predictive value, and positive predictive value. Finally, we examined agreement (K) between Proxy ID and blinded expert assessor categorization of proxy- vs non-proxy SMs.

We defined extent of predicted (informal) proxy use as proportion of PPSMs divided by total SMs for a given patient. To select thresholds to compare characteristics of patients according to extent of total PPSMs, we examined differences in patient age and comorbidity score (Charlson index) separately for patients, exploring each 10% incremental change in frequency of PPSM up to 100%. Based on these results, we constructed a three-level analysis of patient groups with PPSM frequency as follows: (1) "none" (0%), (2) "low" (≥ 0–50%), or (3) "high" (≥ 50–100%). Descriptive statistics, chi-square, and *t* tests were used to compare demographic and clinical characteristics of the 3 groups. To identify variables independently associated with any predicted proxy use, we employed a multivariate logistic regression. Covariates were included in the model based on strength of association and significance ($p < 0.05$) in bivariate analyses. We performed all statistical analyses using SAS.[27]

RESULTS

Registered Proxy-Authored SMs and PPSMs

N-grams were classified into four features: positive unigrams and bi-grams that were associated with registered proxy SMs and negative unigrams and bi-grams that were associated with SMs not registered as proxy. From registered proxy SMs, key unigrams identified by the algorithm included third person pronouns (e.g., he, she, her, his) and specific nouns (e.g., mom, dad, and mother). Key bi-grams included "on behalf," "sent by," and "this message." The MANOVA analysis demonstrated that all n-grams showed significant differences ($p < 0.001$) between the registered and presumed patient-authored SMs. The resultant DFA model reported a classification accuracy for ProxyID of 91.1%, which was significantly higher ($p < 0.001$) than what would be expected by chance.

The sample of 9856 patients generated 375,944 SMs, with mean of 59.92 SMs per patient. The typical patient observation period for SMs was 5.27 ± 2.97 years. Less than 1% of SMs ($n = 2707$) was authored by registered proxies, involving 223 patients (2.3%) (Fig. 1). ProxyID had moderate agreement with blinded expert categorization ($K = 0.58$), with excellent ability to exclude non-proxy SMs (sensitivity 0.93 and NPV 0.95), but less robust specificity (0.70) and PPV (0.64). ProxyID identified 283,582 PPSMs involving 4500 patients (45.7%). In total, 4723 patients (47.9%) had at least one registered proxy SM or PPSM.

Distribution of Extent of Predicted Proxy SM Use

Patients with greater percent PPSMs were significantly older and had more comorbidities. This difference appeared to attenuate when comparing those with > 50% PPSMs to those with < 50% PPSMs. To optimize sample size while attending to these thresholds, we selected a comparison cutoff of 50% to divide patients into "high" (≥ 50–100%) and "low" (≥ 0–50%) percent PPSM groups. Based on ProxyID, we estimated that 40.2% of patients ($n = 3967$) had low frequency PPSMs and 533 patients (5.4%) had high frequency PPSMs. Patients with low frequency PPSMs sent more SMs overall (mean 67.2, SD 58.5) than those with high frequency PPSMs (mean 31.6, SD 47.8).

Patient Sociodemographic Characteristics Associated with Registered Proxy SMs and PPSMs

The table shows that, in comparison to patients with low percent vs no PPSMs, patients predicted as having high percent PPSMs were older (66.5 vs 57.4 vs 56.2 years, $p < 0.001$), more likely to be women (70.4% vs 45.4% vs 45.9%, $p < 0.001$), and non-white (80.7% vs 66.8% vs 72.4%, $p < 0.05$). Those with high percent PPSMs also had higher rates of LEP (16.1% vs 3.2% vs 3.5%, $p < 0.001$) and lower HL scores (3.83 vs 4.43 vs 4.44, $p < 0.001$). Patients with any registered proxy SMs ($n = 223$) significantly differed from those with any PPSMs in terms of age (62.5 vs 58.4 years, $p < 0.05$), race (non-white 79.8% vs 68.5%, $p < 0.05$), but not gender (women 43.0% vs 48.4%, $p > 0.05$) (Table 1).

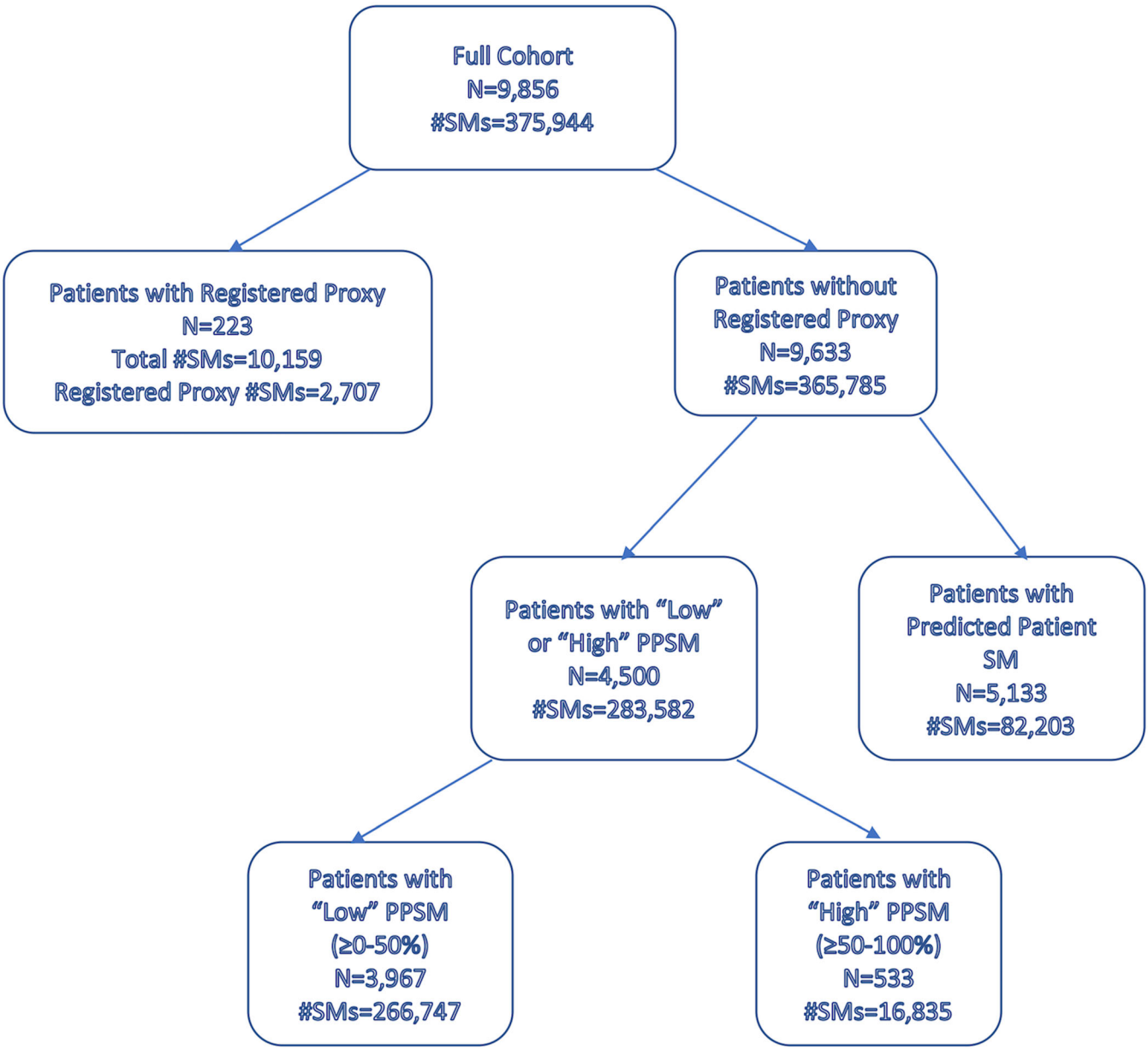

Fig. 1 Patient cohort and SM use by type. The cohort of patients and SM use by type (predicted patient, registered proxy, PPSM).

Patient Clinical Characteristics Associated with PPSM Use

The table shows that patients with high percent PPSMs had more comorbidities (Charlson index 3.78 vs 2.35 vs 2.18, $p < 0.05$) and more frequent healthcare encounters over the past year, including outpatient visits (13.57 vs 10.21 vs 8.88, $p < 0.05$), ED visits, and hospitalizations. Insulin use was more prevalent (38.5% vs 31.9% vs 30.1%, $p < 0.05$) and HbA1c lower (7.40% vs 7.49% vs 7.55%, $p < 0.05$). Patients with high percent PPSMs also had less diabetes knowledge (1.63 vs 2.63 vs 2.59, $p < 0.05$). (Table 1) Patients with any registered proxy SMs did not differ significantly from those with any PPSMs in terms of healthcare utilization, HbA1c, or duration of diabetes, but did differ in Charlson index (3.02 vs 2.52, $p < 0.05$) (Table 2).

Multivariable Regression

The regression model demonstrated that having any PPSMs was independently associated with age 70–79 (OR 1.67, CI 1.41–1.97) vs. age 21–49, non-white race (OR 1.24, CI 1.13–1.37), no educational degree (OR 1.39, CI 1.19–1.63) vs. some college or higher, duration of diabetes ≥ 20 years (adjusted OR 1.45, CI 1.23–1.71) vs. 0–4 years, and Charlson index ≥ 3 vs. 1 (adjusted OR 1.33, CI 1.19–1.49) (Table 2).

DISCUSSION

In a large, ethnically, and socioeconomically diverse sample of patients with diabetes engaged in portal use, we estimated that nearly half of patients likely communicated

Table 1 Characteristics of Patients Predicted to Be Using Proxies for SM Communication

	Total[a] ($n = 9633$)	Predicted proxy SM use: "none," "low," or "high" (0%, ≥0–50%, ≥50–100%)[b]			p value
		"None" ($n = 5133$)	"Low" ($n = 3967$)	"High" ($n = 533$)	
Age (mean, SD)	57.2 ± 10.1	56.2 ± 9.8	57.4 ± 10.1	66.5 ± 8.1	< 0.001
Women	4531 (47.0)	2354 (45.9)	1802 (45.4)	375 (70.4)	< 0.001
Race					< 0.001
White	2833 (29.4)	1415 (27.6)	1315 (33.1)	103 (19.3)	
Black	1403 (14.6)	776 (15.1)	567 (14.3)	60 (11.3)	
Hispanic	1397 (14.5)	712 (13.9)	556 (14.0)	129 (24.2)	
Asian	2944 (30.6)	1680 (32.7)	1101 (27.8)	163 (30.6)	
Other	1053 (10.9)	548 (10.7)	427 (10.8)	78 (14.6)	
Unknown	3 (0.0)	2 (0.0)	1 (0.0)	0 (0.0)	
Income					< 0.001
< $50,000	2865 (34.5)	1432 (32.1)	1143 (33.1)	290 (71.8)	
$50,000–80,000	2335 (28.1)	1312 (29.4)	960 (27.8)	63 (15.6)	
≥ $80,000	3115 (37.5)	1714 (38.4)	1350 (39.1)	51 (12.6)	
Education					< 0.001
No degree	997 (10.5)	438 (8.7)	395 (10.1)	164 (31.6)	
GED/high school	2483 (26.2)	1244 (24.6)	1060 (27.1)	179 (34.5)	
Some college or more	6012 (63.3)	3375 (66.7)	2461 (62.8)	176 (33.9)	
Limited English proficiency[b]	335 (4.0)	155 (3.5)	111 (3.2)	69 (16.1)	< 0.001
Health literacy score (mean, SD)[c]	4.41 ± 0.73	4.44 ± 0.70	4.43 ± 0.72	3.83 ± 0.99	< 0.001
Comorbidity score (Charlson index)	2.34 ± 1.72	2.18 ± 1.62	2.35 ± 1.67	3.78 ± 2.24	< 0.001
Number of chronic medications (past year)	5.49 ± 3.18	5.16 ± 3.09	5.68 ± 3.15	7.28 ± 3.62	< 0.001
Insulin use (past year)	3015 (31.3)	1545 (30.1)	1265 (31.9)	205 (38.5)	0.001
Dementia[d]	80 (0.8)	24 (0.5)	32 (0.8)	24 (4.5)	< 0.001
HbA1c	7.52 ± 1.53	7.55 ± 1.53	7.49 ± 1.54	7.40 ± 1.53	0.009
Diabetes duration (mean, years)	9.81 ± 8.47	9.21 ± 8.06	10.20 ± 8.71	12.71 ± 9.72	< 0.001
Diabetes knowledge[e] (range 0–5)	2.55 ± 1.81	2.59 ± 1.82	2.63 ± 1.78	1.63 ± 1.73	< 0.001
Severe hypoglycemic episodes (past year)[f]					< 0.001
≥ 12	26 (0.3)	8 (0.2)	14 (0.4)	4 (1.1)	
7–11	33 (0.4)	17 (0.4)	11 (0.3)	5 (1.4)	
4–6	101 (1.3)	62 (1.5)	37 (1.1)	2 (0.6)	
1–3	604 (7.6)	295 (7.0)	274 (8.2)	35 (9.8)	
0 or none	7159 (90.4)	3852 (91.0)	2995 (89.9)	312 (87.2)	
No. outpatient visits (past year)	9.69 ± 9.91	8.88 ± 9.51	10.21 ± 10.0	13.57 ± 11.5	< 0.001
No. ED visits (past year)	0.43 ± 1.05	0.35 ± 0.94	0.44 ± 1.04	1.09 ± 1.69	< 0.001
No. inpatient visits (past year)	0.20 ± 0.65	0.15 ± 0.56	0.21 ± 0.67	0.53 ± 1.10	< 0.001

[a]*Patients with SMs authored by registered proxy users (n = 223) are not included*
[b]*Frequency of proxy use was categorized as "none," "low," and "high" (0%, 0–50%, ≥ 50%, respectively), of a patient's total SMs that were either predicted by the NLP-proxy algorithm as likely written by a proxy or were written by a registered proxy*
[c]*DISTANCE Q26. Respondents were asked "How often do you have difficulty understanding or speaking English? Responses were dichotomized as Limited English Proficient (LEP) ("Always," "Often," and "Sometimes") and English Proficient ("Rarely" and "Never")*
[d]*The Health Literacy scores were computed by taking the mean of four self-reported HL measures (DISTANCE Q113-116)*
[e]*ICD 9 medical chart diagnosis of dementia*
[f]*Respondents were asked 5 questions regarding basic diabetes knowledge. Each correct answer received 1 pt. for total score ranging 0 to 5 (DISTANCE Q118-122)*
[g]*Respondents were asked: "In the past year, how many times have you had a severe low blood sugar reaction such as passing out or needing help to treat the reaction? (DISTANCE Q170)*

with their primary care providers via proxies, e.g., family members or other caregivers. The majority of proxy communication was informal (non-registered). We created ProxyID, a novel computational linguistics based algorithm that predicted informal proxy use with moderate accuracy, allowing us to characterize patients reliant on proxies for SM. We found that patients with high percent PPSMs compared to those with low percent or no PPSMs, were more likely to be older, non-white, have higher LEP rates and limited HL, more comorbidities and higher healthcare utilization.

Increasingly, healthcare systems enable portal features that allow access to proxies who communicate with providers on patients' behalf.[9, 28] Yet, the extent of proxy portal use among patients is unclear. A recent study of 20 large healthcare systems found that, while all offered proxy portal access, none reported the number of registered portal users with authorized proxies.[28] We found that only 2.3% of patients had registered proxy-authored SMs, while a much larger percentage (45.7%) had SMs predicted as authored by informal proxies. Other studies report 25–50% of proxies use portals informally.[10, 11] These prior estimates relied on self-reports from proxies who may be reluctant to disclose unauthorized use, thereby underestimating the true extent of informal proxy use.

ProxyID allowed us to provide an objective estimate of prevalence of proxy use in our sample. While we detected PPSMs among nearly half of patients, relatively few patients (5.5%) used proxies more than half the time, suggesting a spectrum of reliance on caregivers for secure messaging. A study exploring VA patients' preferences regarding shared access of their patient portal revealed greater patient interest in proxy access to appointment

Table 2 Logistic Regression Analysis for Likelihood of Patients Using a Proxy for SM

Predictor	OR (CI)
Men	0.944 (0.864–1.032)
Age (vs 21–49 years)	
50–59 years	1.036 (0.921–1.165)
60–69 years	1.130 (0.995–1.283)
70–79 years	1.667 (1.410–1.970)*
Non-White	1.243 (1.129–1.369)*
Education (vs some college or more)	
No degree	1.390 (1.188–1.626)*
GED/high school	1.194 (1.079–1.321)*
Insulin use (past year)	0.984 (0.886–1.093)
Limited English proficiency	1.144 (0.908–1.442)
Diabetes duration (vs 0–4 years)	
5–9 years	1.074 (0.960–1.202)
10–14 years	1.118 (0.977–1.280)
15–19 years	1.192 (1.000–1.420)
≥20 years	1.449 (1.226–1.713)*
Charlson index (vs 1)	
2	1.226 (1.098–1.369)*
≥3	1.332 (1.189–1.491)*

scheduling and prescription refills than in communication with providers.[29] The VA patient sample may not have included the same degree of low HL and LEP as our sample, factors which appear associated with greater need for proxy communication.

Patients with high percent PPSMs were more likely to be women and were on average nearly 10 years older than patients with low percent or no PPSMs. This is not surprising given the barriers to use of patient portals among older populations.[30–32] Additionally, women comprise a majority of elderly care recipients and the recent Health Information National Trends Survey found women engage more than men in health information seeking behaviors, particularly through digital tools.[1, 33] Patients with the highest proportion of PPSMs were also more likely to be non-white, have lower income, and less education. While prior work has demonstrated racial/ethnic minorities and those with lower HL less frequently register on patient portals, there has not been a consistent disparity observed in portal engagement once enrolled.[32, 34, 35] Our findings suggest that racial/ethnic minorities and those with low HL or LEP, who overcome initial barriers to portal enrollment, may be using proxies as one means to engage. Our findings also support previous research that caregivers tend to be interested in using technology to support their caregiving activities.[36]

In our study of patients with diabetes engaged in portal use, we found that high predicted proxy users had clinical characteristics that suggest more complex care needs, such as greater prevalence of insulin use, more comorbidities and higher healthcare utilization. We are not aware of previous research to determine how beneficial proxy SM use is to these patients or caregivers. Prior work has shown that patients with diabetes who actively engage in SM communication with providers via portals have greater likelihood of achieving glycemic control when compared to those who engage more passively or do not engage at all.[37, 38]

Our study has several limitations. Though novel, ProxyID has not been externally validated. Linguistic algorithms have been applied to more structured writing samples as opposed to email messages. While ProxyID demonstrated excellent discriminatory ability in excluding non-proxy messages, the lower specificity suggests that we likely misclassified some patient messages as PPSMs. This may have led to overestimating the number of SMs written by proxies, and inflating the number of patients who used a proxy at least once. In contrast, some proxies may have avoided language within SM text that ProxyID could detect (e.g., "I am writing on behalf of …"). The prevalence of these concealed proxies is unknown; our estimates of informal proxy use could be an underestimate. Our sample was drawn from one healthcare system and results may not be generalizable to other study populations. However, this is a large integrated healthcare system with advanced and frequent portal use among a socioeconomically and ethnically diverse, insured population, with incomes that reflect the US populations' except at the extremes of income. In addition, our findings focus on older adults and are not generalizable to pediatric populations, where parent proxies play a significant role. Finally, examining the content of SMs was beyond the scope of this study; future work may provide insight as to variability in and value of proxy use among patients with diabetes.

Despite these limitations, to our knowledge, this is the first study to employ computational linguistics to estimate patients' use of proxies to communicate on their behalf with physicians via SM. We found that proxy SM use, as predicted by the ProxyID algorithm, appears to be common in patients with diabetes, and that clinically and socioeconomically vulnerable patients—including those with communication barriers—are more likely to use SM via proxies. Our study is particularly timely, given the rapid expansion of patient portals across healthcare systems, the potential of portals to improve access to and quality of healthcare for vulnerable populations, and the value of portal-based communication for caregivers. Future work should explore the types of inquiries proxies make on behalf of patients, and how these inquiries are responded to by providers. In addition, determining whether proxy SM use provides health benefits to patients and/or caregivers may accelerate efforts to increase the prevalence of proxy SM use among those who stand to benefit the most, as well as motivate health system changes to better integrate and accommodate the needs of caregivers.

Acknowledgments: *Danielle McNamara is the founder and owner of Adaptive Literacy Technologies. Funding from the company was not used to underwrite this research and the company will not benefit from this research.*

Corresponding Author: *Wagahta Semere, MD, MHS; Division of General Internal MedicineUniversity of California, San Francisco, San Francisco, CA, USA (e-mail: wagahta.semere@ucsf.edu).*

Author Contributions *Study concept and design: S, C, K, S; acquisition of subjects and/or data: C, K, L, S; analysis and interpretation of data: S, C, K, L, L, S; and preparation of manuscript: S, C, K, L, L, M, B, R, S*

Funding *This work was part of a larger parent study, ECLIPPSE, funded by the NIH/NLM (R01 LM12355). This work was also supported by "The HMO Research Network (HMORN)–UCSF Center for Diabetes Translational Research (CDTR)," funded by the NIH/NIDDK and Kaiser Foundation Research Institute (P30 DK092924).*

Compliance with Ethical Standards:

Conflict of Interest: *The authors declare that they do not have a conflict of interest.*

REFERENCES

1. NAC and AARP Public Policy Institute. Research Report. Caregiving in the U.S. Available at: https://www.aarp.org/content/dam/aarp/ppi/2015/caregiving-in-the-united-states-2015-report-revised.pdf. Accessed June 20, 2019.
2. The National Academies of Sciences, Engineering, and Medicine. Families Caring for an Aging America. Available at: http://www.nationalacademies.org/hmd/Reports/2016/families-caring-for-an-aging-america.aspx. Accessed June 20, 2019.
3. **Ho SC**, **Chan A**, **Woo J**, et al. Impact of caregiving on health and quality of life: a comparative population-based study of caregivers for elderly persons and noncaregivers. J Gerontol A Biol Sci Med Sci. 2009;64:873–879.
4. **Bauer JM**, **Sousa-Poza A**. Impacts of informal caregiving on caregiver employment, health, and family. J Popul Ageing. 2015;8(3):113–145.
5. **Sarkar U**, **Bates DW**. Care partners and online patient portals. JAMA. 2014;311(4):357–8.
6. **Wolff JL**, **Darer JD**, **Larsen KL**. Family Caregivers and Consumer Health Information Technology. J Gen Intern Med. 2016;31(1):117–121.
7. **Zulman DM**, **Piette JD**, **Jenchura EC**, **Asch SM**, **Rosland AM**. Facilitating out-of-home caregiving through health information technology: survey of informal caregivers' current practices, interests, and perceived barriers. J Med Internet Res. 2013;15(7):123
8. **Zulman DM**, **Nazi KM**, **Turvey CL**, **Wagner TH**, **Woods SS**, **An LC**. Patient interest in sharing personal health record information: A web-based survey. Ann Intern Med. 2011 Dec 20;155(12):805–10.
9. The Office of the National Coordinator for Health Information Technology. Patient Engagement Playbook. Chapter 3: Allow Access for Caregivers. Available at: https://www.healthit.gov/playbook/pe/chapter-3/. Accessed June 20, 2019.
10. **Reed ME**, **Huang J**, **Brand R**, et al. Communicating Through a Patient Portal to Engage Family Care Partners. JAMA Intern Med. 2018;178(1)142–144.
11. **Wolff JL**, **Berger A**, **Clarke D**, et al. Patients, care partners, and shared access to the patient portal: online practices at an integrated health system. J Am Med Inform Assoc. 2016 Nov 1;23(1):1150–1158.
12. **Wolff JL**, **Darer JD**, **Berger A**, et al. Inviting patients and care partners to read doctors' notes: OpenNotes and shared access to electronic medical records. J Am Med Inform Assoc. 2017 Apr 1;24(e1):e166-e172.
13. **Aikens JE**, **Trivedi R**, **Heapy A**, **Pfeiffer PN**, **Piette JD**. Potential Impact of Incorporating a Patient-Selected Support Person into mHealth for Depression. J Gen Intern Med. 2015 Jun;30(6):797–803.
14. **Price-Haywood EG**, **Harden-Barrios J**, **Ulep R**, **Luo Q**. eHealth Literacy: Patient Engagement in Identifying Strategies to Encourage Use of Patient Portals Among Older Adults. Popul Health Manag. 2017;20(6):486–494.
15. **Sarkar Urmimala**, **Karter Andrew J**, **Liu Jennifer Y**, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc. 2011 May 1;18(3):318–21.
16. **Irizarry T**, **Shoemake J**, **Nilsen ML**, **Czaja S**, **Beach S**, **DeVito Dabbs A**. Patient Portals as a Tool for Health Care Engagement: A Mixed-Method Study of Older Adults with Varying Levels of Health Literacy and Prior Patient Portal Use. J Med Internet Res. 2017 Mar 30;19(3)e99.
17. **Gordon NP**, **Hornbrook MC**. Differences in Access to and Preferences for Using Patient Portals and Other eHealth Technologies Based on Race, Ethnicity, and Age: A Database and Survey Study of Seniors in a Large Health Plan. J Med Internet Res. 2016 Mar 4;18(3):e50.
18. **Krist Alex H**, **Woolf Steven H**, **Bello Ghalib A**, et al. Engaging primary care patients to use a patient-centered personal health record. Ann Fam Med. 2014;12(5):418–26.
19. **Nguyen A**, **Mosadeghi S**, **Almario CV**. Persistent digital divide in access to and use of the Internet as a resource for health information: Results from a California population-based study. Int J Med Inform. 2017 Jul;103:49–54.
20. **Schillinger D**, **McNamara D**, **Crossley S**, et al. The Next Frontier in Communication and the ECLIPPSE Study: Bridging the Linguistic Divide in Secure Messaging. J Diabetes Res. 2017;2017:1348242.
21. **Moffet HH**, **Adler N**, **Schillinger D**, et al. 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. 2009 Feb;38(1): 38–47.
22. **Sarkar U**, **Karter AJ**, **Liu JY**, et al. The Literacy Divide: Health Literacy and the Use of an Internet-Based Patient Portal in an Integrated Health System—Results from the Diabetes Study of Northern California (DISTANCE). J Health Commun. 2010;15(Suppl 2):183–196.
23. **Crossley SA**, **Allen LK**, **Kyle K**, **McNamara DS**. Analyzing Discourse Processing Using the Simple Natural Language Processing Tool (SiNLP). Discourse Processes. 2014;51(5–6): 511–534.
24. **McEnery T**, **Hardie A**. Corpus Linguistics: Method, Theory and Practice. International Journal of Corpus Linguistics. 2013 Jan;18(2):290–294.
25. United States Census Bureau. Language Use. Available at: https://www.census.gov/topics/population/language-use/about.html. Accessed June 20, 2019.
26. **Sarkar U**, **Karter AJ**, **Liu JY**, et al. Validation of self-reported health literacy questions among diverse English and Spanish-speaking populations. J Health Commun. 2010; 15(Suppl 2): 183–196.
27. SAS. Version 9.4. Cary, NC: SAS Institute Inc., 2014.
28. **Wolff JL**, **Kim VS**, **Mintz S**, **Stametz R**, **Griffin JM**. An environmental scan of shared access to patient portals. J Am Med Inform Assoc. 2018; 25(4): 408–412.
29. **Zulman DM**, **Nazi KM**, **Turvey CL**, **Wagner TH**, **Woods SS**, **An LC**. Patient Interest in Sharing Personal Health Record Information: A Web-Based Survey. Ann Intern Med. 2011;155:805–810.
30. **Tieu L**, **Sarkar U**, **Schillinger D**, et al. Barriers and Facilitators to Online Portal Use Among Patients and Caregivers in a Safety Net Health Care System: A Qualitative Study. J Med Internet Res. 2015;17(12):e275.
31. **Wu Y-H**, **Damnée S**, **Kerhervé H**, **Ware C**, **Rigaud A-S**. Bridging the digital divide in older adults: a study from an initiative to inform older adults about new technologies. Clinical Interventions in Aging. 2015;10:193–201.
32. **Kontos E**, **Blake KD**, **Chou W-YS**, **Prestin A**. Predictors of eHealth Usage: Insights on The Digital Divide from the Health Information National Trends Survey 2012. Eysenbach G, ed. J Med Internet Res. 2014;16(7):e172.
33. National Cancer Institute. HINTS 2017: Women and Health Information Seeking. Available at: http://www.nationalpartnership.org/research-library/health-care/hints-2017-women-and-health-info-seeking.pdf. Accessed June 20, 2019.
34. **Goel MS**, **Brown TL**, **Williams A**, **Hasnain-Wynia R**, **Thompson JA**, **Baker DW**. Disparities in Enrollment and Use of an Electronic Patient Portal. J Gen Intern Med. 2011;26(10):1112–1116. doi:https://doi.org/10.1007/s11606-011-1728-3.
35. **Smith SG**, **O'Conor R**, **Aitken W**, **Curtis LM**, **Wolf MS**, **Goel MS**. Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. J Am Med Inform Assoc. 2015;22(4):888–895.
36. American Association of Retired Persons. Caregivers and Technology: What they Want and Need. Available at:https://www.aarp.org/content/dam/aarp/research/surveys_statistics/ltc/2018/caregivers-technology-needs.doi.10.26419-2Fres.00191.002.pdf . Accessed June 20, 2019.
37. **Shimada SL**, **Allison JJ**, **Rosen AK**, **Feng H**, **Houston TK**. Sustained Use of Patient Portal Features and Improvements in Diabetes Physiological Measures. Eysenbach G, ed. J Med Internet Res. 2016;18(7):e179.
38. **Devkota B**, **Salas J**, **Sayavong S**, **Scherrer JF**. Use of an Online Patient Portal and Glucose Control in Primary Care Patients with Diabetes. Popul Health Manag. 2016; 19(2): 125–131.