Physician and Patient Perspectives on Treatment Success Captured in Clinical Progress Notes

Sarah B. Floyd, Jordyn Chandler Sutton, Marvin Okon, Mary McCarthy, Liza Fisher, Benjamin Judkins, Zachary Cole Reynolds, Ann Blair Kennedy

Submitted to: Journal of Participatory Medicine
on: May 06, 2024

Table of Contents

Physician and Patient Perspectives on Treatment Success Captured in Clinical Progress Notes

Sarah B. Floyd[1] MPH, PhD; Jordyn Chandler Sutton[1] BS; Marvin Okon[1] MD; Mary McCarthy[2]; Liza Fisher[2]; Benjamin Judkins[3] MD; Zachary Cole Reynolds[4] MD; Ann Blair Kennedy[5] DrPH

[1]Department of Public Health Sciences Clemson University Clemson US
[2]Patient Engagement Studio, University of South Carolina Greenville US
[3]Prisma Health Greenville US
[4]Prisma Health Greenville, SC US
[5]University of South Carolina School of Medicine Greenville US

Corresponding Author:
Sarah B. Floyd MPH, PhD
Department of Public Health Sciences
Clemson University
503 Edwards Hall
Clemson
US

Abstract

Background: Clinical progress notes are generated for patients during each healthcare encounter and include information on current disease symptoms and rehabilitation progress. These notes provide a rich data source for the application of natural language processing (NLP) methods to identify patients experiencing successful treatment outcomes. However, to use NLP methods to develop an outcome measure that is reflective of both physician and patient perspectives, there needs to be an agreed upon set of patient experiences that reflect treatment success.

Objective: The objective of this paper was to assess if physicians and patients agree on whether patient experiences captured in clinical progress notes reflect a successful patient outcome following orthopaedic treatment.

Methods: We performed a cross-sectional analysis of a subset of clinical notes for patients presenting to a Level-1 Trauma Center Regional Health System for follow-up for an acute Proximal Humerus Fracture (PHF). Physicians and patients reviewed a sample of 100 clinical notes and labeled each note as reflecting one of four treatment outcome labels (treatment success, improvement, deterioration, or failure). The four outcome levels were then aggregated into a binary classifier indicating treatment success or failure, and agreement between physicians and patient labels were assessed. Descriptive analyses were performed to report characteristics of the sample. Cohen's Kappa statistics were used to assess the degree of agreement between physician and patient evaluators.

Results: The average age of the patients in the sample was 67 years of age (SD= 13.6) and 82% of the notes came from female patients. Patients were primarily white (91%) and had Medicare insurance coverage (65%). The note sample came from fracture-related encounters ranging from the second to the 10th encounter after the PHF visit. There were no significant differences in patient or visit characteristics across concordant and discordant notes. Physician and patient evaluators exhibited a fair level of agreement in what they deemed as treatment success based on a Cohen's Kappa of 0.32 ((95% CI, 0.10 to .55), p = .01).

Conclusions: Findings suggest that physicians and patients only demonstrated fair agreement when interpreting whether a patient had achieved treatment success following treatment for an acute shoulder fracture. Further research is needed to examine how different perceptions of treatment success may influence the development of NLP-based outcome measures.

(JMIR Preprints 06/05/2024:60263)
DOI: https://doi.org/10.2196/preprints.60263

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Physician and Patient Perspectives on Treatment Success Captured in Clinical Progress Notes

Sarah Floyd, PhD, MPH[1]
Jordyn Sutton, BS[1]
Marvin Okon, MD[1]
Mary McCarthy, BS[3]
Liza Fisher, MBA[3]
Benjamin L. Judkins, MD[4]
Zachary Cole Reynolds, MD[4]
Ann Blair Kennedy, LMT, BCTMB, DrPH[2, 3]

1 Department of Public Health Sciences, Clemson University, Clemson, SC
2 Patient Engagement Studio, University of South Carolina, Greenville, SC
3 University of South Carolina School of Medicine Greenville, Department of Biomedical Sciences, Greenville, SC
4 Prisma Health, Department of Orthopaedic Surgery, Greenville, SC

Corresponding Author:
Sarah Floyd, PhD
Department of Public Health Sciences
College of Behavioral, Social and Health Sciences
Clemson University
Email: sbf@clemson.edu

Funding: AHRQ R03 Funding (R03HS029060)

Abstract

Background: Clinical progress notes are generated during each healthcare encounter and include information on current disease symptoms and rehabilitation progress. These notes provide a rich data source for the application of natural language processing (NLP) methods to identify patients experiencing successful treatment outcomes. However, to use NLP methods to develop an outcome measure that is reflective of both physician and patient perspectives, there needs to be an agreed upon set of patient experiences that reflect treatment success.
Objective: The objective of this paper was to assess if physicians and patients agree on whether patient experiences captured in clinical progress notes reflect a successful patient outcome following orthopaedic treatment.
Methods: We performed a cross-sectional analysis of a subset of clinical notes for patients presenting to a Level-1 Trauma Center and Regional Health System for follow-up for an acute Proximal Humerus Fracture (PHF). Physicians and patients reviewed a sample of 100 clinical notes and labeled each note as reflecting one of four treatment outcome labels (treatment success, improvement, deterioration, or failure). The four outcome levels were then aggregated into a binary classifier indicating treatment success or failure, and agreement between physicians and patient labels were assessed. Descriptive analyses were performed to report characteristics of the sample. Cohen's Kappa statistics were used to assess the degree of agreement between physician and patient evaluators.
Results: The average age of the patients in the sample was 67 years of age (SD= 13.6) and 82% of the notes came from female patients. Patients were primarily white (91%) and had Medicare insurance coverage (65%). The note sample came from fracture-related encounters ranging from the second to the 10^{th} encounter after the index PHF visit. There were no significant differences in patient or visit characteristics across concordant and discordant notes. Physician and patient evaluators exhibited a fair level of agreement in what they deemed as treatment success based on a Cohen's Kappa of 0.32 ((95% CI, 0.10 to .55), p = .01).
Conclusions: Findings suggest that physicians and patients only demonstrated fair agreement when interpreting whether a patient had achieved treatment success following treatment for a PHF. Further research is needed to examine how different perceptions of treatment success may influence the development of NLP-based outcome measures.
Key words: Patient-reported outcomes; Acute proximal humerus fracture; Patient involvement; Orthopaedic

Introduction

Widely available outcome measures are needed in orthopaedic medicine to conduct large scale effectiveness studies to develop evidence to enhance shared decision-making and guide treatment decision-making for individual patients.[1-4] Patient-reported outcome measures (PROMs), a first-hand assessment of a patient's physical capabilities and symptoms, are commonly used in orthopaedic research for evaluating the success of orthopaedic care.[5, 6] However, PROM collection in real-world practice is difficult due to challenges with technology integration, data collection disruptions to clinic workflows, language differentials, and low completion rates of PROM questionnaires.[7-9] Thus without widely available patient outcome data to generate evidence on the effectiveness of treatments for given patients, treatment decisions in orthopaedics remain highly discretionary and highly varied.[10, 11] An alternative to patient outcome collection could improve the quality of decision making in orthopaedic medicine.[1, 12]

The electronic health record (EHR) system is the primary tool to document and store records of patient encounters in hospitals and outpatient clinics in the United States.[13-16] Clinical progress notes are generated for each encounter that patients have with their physician or healthcare provider. These contain rich information on current disease symptoms, rehabilitation progress, and unexpected complications.[16-23] The subjective, objective, assessment and plan sections of the clinical progress note reflect the patient's story, told to and interpreted by the physician.[8, 22, 24] Unstructured progress notes produce a record of a patient's history, physical findings, medical reasoning, and patient care and [8, 24-26] reveal distinct trajectories of patient outcomes after treatment.[24] In successful cases, the progress note documents the degree of improvement or relief experienced and reported directly by patients.[23] Conversely, when symptoms have not been resolved, are lingering, or when subsequent complications have arisen, these ongoing patient complaints and persistent treatment utilization are documented in the notes.[27] The content within the progress note highlights symptoms and outcome dimensions which are valued by patients and contribute to a patient's individual definition of success.[18]

Applications of natural language processing (NLP) methods in healthcare are rapidly expanding. [17, 19, 28-30] NLP methods can be used to identify phrases of treatment success or failure captured in progress notes and discern an indicator of treatment success for patients.[31] However, the use of NLP methods to develop an outcome measure that is reflective of both physician and patient perspectives of success requires an agreed upon definition of treatment success that can be used to train the NLP model.[32, 33] The degree of pain relief, joint function, and ability to engage in activities of daily living and physical activities, which are key dimensions of treatment success, are highly patient-dependent.[34, 35] On the other end, objective events like surgical infections, prosthetic failure or avascular necrosis are commonly agreed upon as poor patient outcomes for all orthopaedic patients.[27]

The objective of this paper was to assess if physician and patient evaluators agree on whether patient experiences captured in clinical progress notes reflect a successful patient outcome following orthopaedic treatment. As the use of NLP methods in healthcare continues to expand, it is important to evaluate how models are trained and specifically whose voice is used as the gold standard for model training. To date, no studies have

assessed whether there is agreement between physicians and patients about whether the progress note reflects treatment success or failure. In this paper we describe our process for patient involvement in the development of a NLP text classifier model and assess if physicians and patients agree on whether examples of clinical progress notes reflect a successful patient outcome following orthopaedic treatment for an acute proximal humerus fracture (PHF). Shoulder fracture was the focus of this work because it remains a challenging clinical condition to treat, [36-44] and one that could greatly benefit from widely available outcome data to guide clinical decision-making.[36-38, 42, 45]

Methods

Study Sample

This was a cross-sectional analysis of a subset of progress notes from a larger study. The study included adult patients presenting to a Level-1 Trauma Center Regional Health System for an acute Proximal Humerus Fracture (PHF) between January 1, 2019 and December 31, 2021. The first visit for a patient for PHF during the study period was defined as the index visit. We then identified all health system encounters (hospital encounters, office visits, etc.) with a diagnosis of PHF or shoulder pain from the index PHF visit to 365 days after the index PHF visit. Of those encounters, we took the progress note from the last office visit for PHF-related care (ICD10: S42.2XXX) or shoulder pain (ICD10: M45.2XXX) to occur before 365 days post-index. This resulted in one note per person. Patients were excluded from the study if they were less than 18 years of age, did not have at least one office visit with a diagnosis of PHF or shoulder pain that occurred 45 days or more days after the index visit, or if their last office visit was less than 500 characters. A minimum of 45 days after index was used as this is the minimal time needed for healing of a PHF, before which treatment success cannot be assessed. The larger study included a sample of 1,000 patients meeting these inclusion criteria. For this study, a stratified random sample of 100 progress notes was used to assess agreement between physicians and patients on their perceptions of treatment success. This study was approved by the Prisma Health Institutional Review Board (1924627-1).

Treatment Label Development Process

The University of South Carolina Patient Engagement Studio (PES) brings together patients and caregivers, community groups, health system innovators and clinicians, and academic researchers to produce meaningful research and innovation that advances health and research outcomes. The senior author (S.F.) led three PES sessions to engage patients in contributing to the development of a definition of treatment success in orthopaedics. There is no one standard acceptable definition of treatment success, as it varies based upon patient lifestyle and desired goals.[34, 46] However, the general goal of orthopaedic treatment is to restore joint function, minimize pain and maximize quality of life for patients after an injury. [47] Together, the PES members and senior author defined four distinct outcome states, associated definitions, and indicators that described the range of positive and negative outcomes that could be found in progress notes following care for PHF. Figure 1 contains the four distinct outcome states, associated definitions, and indicators.

Note Labeling Process

To assess agreement between physician and patient evaluators on whether note content reflected treatment success, a subset of notes was reviewed by orthopaedic residents and patient evaluators. As part of a related study, orthopaedic residents labeled the full sample of 1,000 notes. From that labeled dataset, a stratified random sample was used to identify 25 notes from each outcome label group for this study, for a total of 100

notes. Each orthopaedic resident received a one-hour training on the note labels and definitions. Residents were instructed to assess the current state of the patient reflected in the note. The orthopaedic residents included 3 male and 1 female, and all had been working as an orthopaedic resident for a minimum of 2 years. An attending orthopaedic surgeon, and the Department of Orthopaedic Surgery Chair, served as the final note evaluator when discordance occurred between residents' labels.

Two members of the PES agreed to participate in this study. Similar to orthopaedic residents, patient evaluators also received a one-hour training on the note labels and definitions. Both patient evaluators reviewed all 100 notes and provided labels. A full cross-sample was used so that agreement between patient evaluators could also be assessed. Both patient evaluators were female and had personal experience with musculoskeletal injuries. The Program Manager of the PES (K.P.) served as the final note evaluator when discordance occurred between patient evaluators. All progress notes were redacted to conceal patient identifiers prior to patient review. REDCap [48, 49] hosted by Prisma Health was used to organize and store physician and patient labels for each note.

Statistical Analysis

The four outcome labels were aggregated into a binary classifier representing treatment success or failure. Success was represented by notes labeled "Treatment Success." The three remaining labels, including "Partial Treatment Success- improvement of condition", "Partial Treatment Success- deterioration of condition" and "Treatment Failure" were grouped into the Treatment failure group. Treatment failure was comprised of all labels with documentation of lingering, symptomatic problems requiring ongoing care.

Agreement between physicians and patients was calculated across binary groups of treatment success and failure. Descriptive analyses were used to assess characteristics of the progress note samples and T-tests and Chi square tests were used to assess differences in concordant and discordant notes. Discordant labels were defined as notes with differing outcome states provided by the respective labelers. Cohen's kappa statistics were used to assess the degree of agreement between patient labels as well as the degree of agreement between physician and patient labels. We used the benchmarks for agreement for categorical data as described by Landis and Koch, where 0.00 to 0.20, 0.21 to 0.40, 0.41 to 0.60, 0.61 to 0.80, and 0.81 to 1.00 indicate poor, fair, moderate, substantial, and almost perfect agreement, respectively.[50] A Bangdiwala agreement chart is presented to display agreement between physician and patient labels.[51] Analyses were performed with SAS (Cary, NC) version 15.2, R studio, and Microsoft Excel.

Results

Progress Note Characteristics

The sample of 100 progress notes for this study came from patients treated across 24 departments and 54 distinct physicians within one regional health system. The average age of the patient was 67 years of age (SD= 13.6) and 82% of the notes came from female patients. Patients were primarily white (91%) and had Medicare insurance coverage (65%). The note sample came from fracture-related encounters ranging from the second to the 10th encounter after the index PHF visit, with a mean length of 145 days after index (range 45 to 362 days). The progress notes text lengths ranged from 981 to 15,297 characters with a mean length of 5,506 characters. There was no significant difference in progress note characteristics across concordant and discordant notes (Table 1).

Agreement between Patients

Both patient evaluators were assigned the full sample of 100 notes to review and label. Of the 100 notes, 34 notes were discordant between patient evaluators. There was a statistically significant level of agreement between the two patient evaluators (Cohens k = 0.41 (95% CI, 0.23 to 0.59), $P < .001$), and the strength of agreement was classified as moderate, according to Landis and Koch.[50]

Agreement between Physicians and Patients

A total of 22 notes were discordant between physicians and patient evaluators. Of the 25 notes labeled as Treatment Success by orthopaedic physicians, 11 notes were also labeled as Treatment Success by patients. The remaining 14 Treatment Success notes were labeled as Treatment Failure by patient evaluators. Of the 75 notes deemed as Treatment Failure, 67 were also labeled as Treatment Failure by patient evaluators. There was a statistically significant level of agreement between orthopaedic physicians and patient evaluators (Cohen's k = 0.32 (95% CI, 0.10 to 0.55), $P = .01$). The strength of agreement between patients and residents was classified as fair, according to Landis and Koch.[50]

Although not the focus of this manuscript, physician agreement was assessed using the larger sample of 1,000 notes. Agreement between physicians was assessed using Fleiss' kappa and agreement between orthopaedic physicians was moderate (Fleiss' k = 0.49 (95% CI, 0.30 to 0.68), $P < 0.05$). [52]

Discussion

The objective of this paper was to assess if physicians and patients agree on whether examples of clinical progress notes reflect a successful patient outcome following orthopaedic treatment. This is an important question to answer for the field of orthopaedic medicine, in addition to needs surrounding the use of NLP on clinical text. The results of this analysis indicate physicians and patient evaluators only demonstrate a fair level of agreement when interpreting if a patient's clinical note reflects treatment success following care for an acute PHF. Further, inter-patient and inter-physician agreement also demonstrated relatively low levels of agreement, signaling that within patients and physician groups, the definition of success is not clearly defined or agreed upon.

We chose to examine this question for patients undergoing PHF treatment and subsequent healing. Shoulder fracture remains a challenging clinical condition to treat, [36-44] and one that could greatly benefit from widely available outcome data to guide clinical decision-making.[36-38, 42, 45] We did not find systematic differences in patient or visit characteristics across notes that were concordant and discordant between physicians and patients. There are treating physician characteristics, such as time in practice, subspeciality, or training, that may result in differences in note documentation signatures[53, 54] that may be associated with higher or lower levels of agreement or disagreement between physicians and patients[55].

We found that physicians and patients only demonstrated a fair level of agreement in their evaluation of treatment success documented in progress notes. It is possible that physicians may be more familiar with clinical note phrasing than patients and that physician familiarity with clinical note documentation may drive differences in note interpretation across groups. [56] To this point, a prior study indicates that common medical phrases used in clinical progress notes may lead to confusion regarding health outcomes due to a difference in health literacy among patients and physicians.[57]

Another plausible explanation for this finding could related to the multitude of outcome dimensions encompassed within in a clinical note (e.g., pain, quality of life, range of motion), along with the varying value assessments assigned to each dimension.[58, 59] Variations in these reported outcomes may impact interpretations of clinical notes and thus, perspectives on treatment success and treatment failure in orthopaedics. [53, 60] Specifically, physicians may have a more clinically based definition of treatment success driven by objective measures such as range of motion. [33] Patients may be more focused on subjective pain and joint function capability as it relates to daily activities and quality of life. [61, 62] It is likely that objective measures may lend themselves to benchmarks and may be easier to assess for all patients, whereas subjective, lifestyle goals and quality of life are harder to identify and assess for individual patients.[62] This thought is supported by prior research which found that physicians and patients have higher levels of agreement for dimensions of physical functioning, compared to the emotional or psychological aspects of functioning.[63-66] Alternatively, it may also be possible that physicians may have different expectations of patient's capabilities following a serious upper extremity injury, such as PHF.[67, 68] Taken together, physicians and patients may find higher levels of agreement in objective outcomes and may experience more disagreement in the areas of subjective outcomes.[66, 69]

Other studies measuring patient and physician agreement post-orthopaedic surgery concluded patients and physicians agree less when the patient had worse health outcomes. [1, 70, 71] These conclusions are not consistent with our study findings. We found that physicians and patients were in agreement for a larger share of the treatment failure notes (89% or 67/75 notes), compared to only 44% of the treatment success notes (11 of 25 notes). It is our belief that treatment failure is more clear cut (e.g. surgical complications, persistent pain, nonunion), whereas treatment success is more patient-specific.[1] Consequently, while it may be easier to recognize when outcomes are unfavorable, pinpointing a positive outcome proves challenging due to the diverse nature of favorable outcomes across patient populations[46, 72]. Unfortunately, there is not one singular definition of patient success, as previous research indicates success depends on a patient's lifestyle and desired goals. [3, 33, 57, 71, 73] In a study assessing patient-physician agreement on management of musculoskeletal injuries and pain associated with those injuries, authors found that patients and physicians prioritize different goals when assessing a patient's treatment outcome.[1, 74]

Limitations and Future Work

Our work has several limitations which should be acknowledged. First, we used a relatively small sample of progress notes from one clinical condition that lacks patient diversity, and our results are highly reflective of the small sample of physicians (n=4) and patient evaluators (n=2) who that completed the labeling. Therefore, our results do not generalize beyond the clinical condition and evaluators we explored in this study. Furthermore, we were unable to assess the characteristics of treating physicians who authored the progress notes. It is possible an imbalance in treating physician characteristics may explain some of the discordance in note labels. Additionally, we worked with resident physicians, who are early in their training and may be less experienced in assessing patient outcomes following care. [75] Also, the way in which we aggregated patient labels may influence the level of agreement we observed. Additionally, physicians often have varying definitions of success for different patients [76]. For example, an elderly patient may prioritize pain reduction and a lower range of motion than a younger more active patient. This dynamic definition of success may have been hard to assess without a complete patient case presentation. Lastly, our agreement results may only reflect orthopaedic physician and patient agreement in defining success and may not generalize to other types of physicians.[77]

Although outside the scope of this work, there are mixed reports of the accuracy and completeness, [63] and quality, [78, 79] of progress note content. However, multiple studies have found that healthcare professionals produce accurate documentation for concrete and overt symptoms, such as range of motion and impaired physical functioning.[80-82] Orthopaedic progress encounters are highly focused on musculoskeletal conditions or injuries and how musculoskeletal conditions affect outcome dimensions such as pain and range of motion and ultimately impact a patient's quality of life.[83, 84] Because of the highly focused nature of orthopaedic encounters, we believe progress note content is a potential source to develop an indicator of treatment success for patients.[19, 59, 85] In future work we plan to expand the orthopaedic note content to include encounters beyond just clinical encounters including physical therapy sessions.

Conclusion

The objective of this study was to assess if physicians and patients agree on whether patient experience captured in clinical progress notes reflect a successful patient outcome following orthopaedic treatment. In performing a cross-sectional analysis of clinical progress notes from an acute follow-up of patients treated for a PHF, we found fair agreement between patients' and physicians' assessments of patient experience reflecting treatment success. Like prior research, these results indicate that patients and physicians do not fully agree on what constitutes treatment success. [1, 55, 86] Our findings emphasize the need to analyze both patient and physician perspectives when determining treatment success. Further research is needed to examine how different perceptions of treatment success may influence outcome development and use in orthopaedic medicine.

Acknowledgements

We want to acknowledge the contributions of the orthopaedic residents that completed the note labeling. We thank Drs. Kyle Jeray, Jahan Threeths, and Claire Krohn. We also thank Katie Parris, the Patient Engagement Studio Program Manager, for her partnership completing the patient labels.

Conflicts of Interest

None Declared.

Abbreviations

PHF: Proximal Humerus Fracture
NLP: Natural Language Processing
PES: Patient Engagement Studio

Figure 1. Treatment Outcome States, Definitions and Indicators Developed by Patient Engagement Studio and Research Team Members

Outcome State	*Definition*	*Example Indicators of Outcome State found in the Clinical Note*
Treatment Success	Treatment success occurs when a patient is able to resume desired activities, has a sufficient range of motion, and is in minimal/mild or no pain. After PHF it is possible for there to be some lingering motion limitations (patient may never return to 100%) or minimal pain, but these issues should not require ongoing treatment or be prohibitive to their desired lifestyle or daily activities.	• Radiographic healing noted on X-ray • Making good progress/improvements with current treatment or stopping treatment • Patient has returned to work or play. • No major complaints documented. • Only follow-up as needed
Improvement of Condition	Improvement occurs when there is a record of some levels of pain or functional problems that are somewhat prohibitive to the patient's desired activities, but improvement is occurring. In these situations, physicians may continue to monitor patients, but do not alter care or treatment courses.	• Radiographic healing or signs of healing occurring. • Moderate loss of function or pain which interferes with desired activities, but no change in treatment. • Ongoing treatment and monitoring progress • Return in 2-6 weeks for repeat X-rays and re-check
Deterioration of Condition	Deterioration occurs when there is a record of some levels of pain or functional problems that are becoming more prohibitive to the patient's desired activities. No real improvements are occurring, and physicians may escalate or alter care or treatment courses.	• Negative radiographic changes observed. • Moderate loss of function or pain which interferes with desired activities requiring a change in treatment. • Initiating or continuing treatment and monitoring progress. • Return in 2-6 weeks for repeat X-rays and re-check.

Treatment Failure	Treatment failure occurs when the patient is experiencing significant pain or limitations and requires subsequent fracture-related care. Failing is occurring when patients are unable to resume desired activities and may include fracture sequelae, complications, or nonunion.	• Ongoing, persistent treatment (injections, surgeries) for symptoms related to PHF. • Unrelenting pain • Surgical complications • Loss of significant motion • Extreme pain • Fracture related sequelae (e.g., avascular necrosis)

Table 1. Patient and Visit Characteristics of the Clinical Progress Note Sample Presented by Patient and Physician Agreement (N=100)

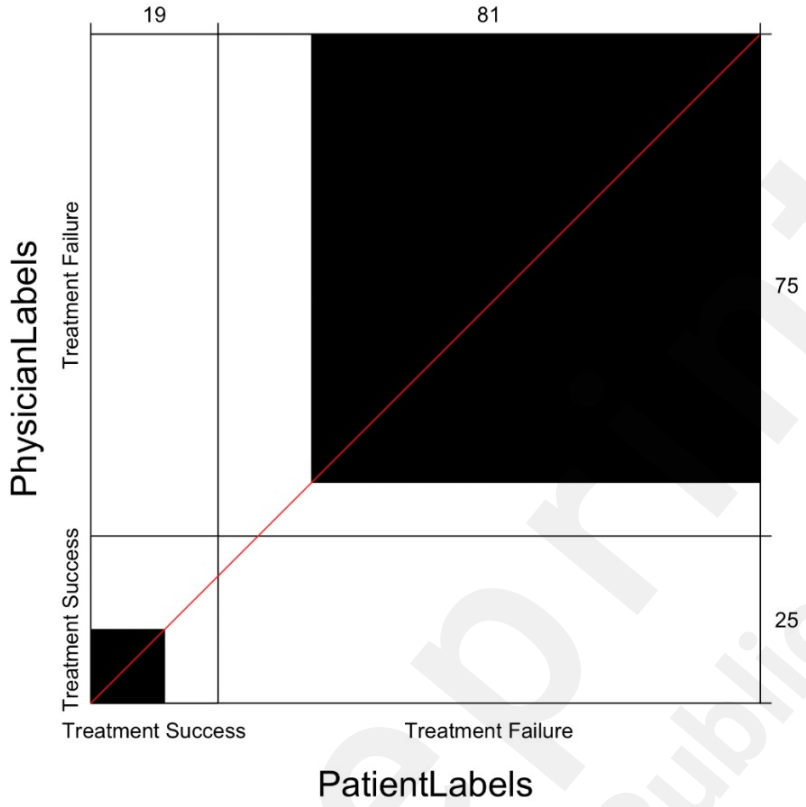
	Total Sample	Concordant Notes	Discordant Notes	*P*-value
	N=100	N=78	N=22	
Patient characteristics				
Patient Age, Mean (SD)	67 (13.6)	66.7 (13.8)	67.9 (13.4)	0.73
Patient Sex, N (%)				0.22
Male	18 (18%)	16 (20%)	2 (9%)	
Female	82 (82%)	62 (79%)	20 (90%)	
Patient Race, N (%)				0.72
White	91 (91%)	71 (91%)	20 (91%)	
Black	5 (5%)	3 (4%)	2 (9%)	
American Indian or Alaskan	1 (1%)	1 (1%)	0 (0%)	
Hispanic	1 (1%)	1 (1%)	0 (0%)	
Unknown	2 (2%)	2 (3%)	0 (0%)	
Insurance Provider				
Medicare	65 (65%)	51 (65%)	14 (64%)	0.44
Medicaid	7 (75)	7 (9%)	0 (0%)	
Private	21 (21%)	15 (19%)	6 (27%)	
Other	7 (7%)	5 (6%)	2 (9%)	
Visit Characteristics				
Days from Index, Mean (SD)	143 (86.9)	145.5 (87.4)	136.0 (86.5)	0.65
PHF-related encounter, Mean (SD)	4.5 (2.2)	4.5 (2.2)	4.7 (2.2)	0.58
Patient treated surgically, N (%)	25 (25%)	21 (27%)	4 (18%)	0.40
Note character length, Mean (SD)	5,506 (3,027)	5,730 (3,165)	4,710 (2,369)	0.16

Table 2. Agreement in Notes Labels Between Physicians and Patients (N=100)

		Patient Rater 2			
Patient Rater 1		Success	Failure	Total	Agreement
	Success	15	4	19	**Moderate** (k= 0.41)*
	Failure	20	61	81	
	Total	35	65	100	
		Patient Labels			
Physician Labels		Success	Failure	Total	Agreement
	Success	11	14	25	**Fair** (k= 0.32)*
	Failure	8	67	75	
	Total	19	81	100	

* Cohens Kappa used to assess agreement. 0.00 to 0.20, 0.21 to 0.40, 0.41 to 0.60, 0.61 to 0.80, and 0.81 to 1.00 indicate poor, fair, moderate, substantial, and almost perfect agreement.[50]

Figure 2. Bangdiwala Agreement Chart[1] for Physician and Patient Note Labels (N=100) [1]



1 Bangidwala chart used to assess agreement between patients and physician's indications of treatment success or treatment failure from analyzed clinical notes. Black boxes indicate overlap of agreement. [51]

1

References

1. Noble, P.C., et al., *Challenges in outcome measurement: discrepancies between patient and provider definitions of success.* Clin Orthop Relat Res, 2013. **471**(11): p. 3437-45.

2. Klifto, K., C. Klifto, and J. Slover, *Current concepts of shared decision making in orthopedic surgery.* Current Reviews in Musculoskeletal Medicine, 2017. **10**(2): p. 253-257.

3. Wilson, C.D. and R.A. Probe, *Shared Decision-making in Orthopaedic Surgery.* J Am Acad Orthop Surg, 2020. **28**(23): p. e1032-e1041.

4. Nathan, K., et al., *Deciding without data: clinical decision-making in pediatric orthopedic surgery.* International Journal for Quality in Health Care, 2020. **32**(10): p. 658-662.

5. Gagnier, J.J., *Patient reported outcomes in orthopaedics.* J Orthop Res, 2017. **35**(10): p. 2098-2108.

6. Lizzio, V.A., M.S. Dekhne, and E.C. Makhni, *Electronic Patient-Reported Outcome Collection Systems in Orthopaedic Clinical Practice.* JBJS Reviews, 2019. **7**(7).

7. Steinbeck, V., et al., *Electronic Patient-Reported Outcome Monitoring to Improve Quality of Life After Joint Replacement: Secondary Analysis of a Randomized Clinical Trial.* JAMA Netw Open, 2023. **6**(9): p. e2331301.

8. Walsh, S.H., *The clinician's perspective on electronic health records and how they can affect patient care.* Bmj, 2004. **328**(7449): p. 1184-7.

9. Quinn, M., et al., *Electronic health records, communication, and data sharing: challenges and opportunities for improving the diagnostic process.* Diagnosis (Berl), 2019. **6**(3): p. 241-248.

10. Floyd, S., et al., *Geographic variation in the treatment of proximal humerus fracture: an update on surgery rates and treatment consensus.* Journal of Orthopaedic Surgery and Research, 2019. **14**.

11. Lisanne Johanna, H.S., et al., *Variation in Nonsurgical Treatment Recommendations for Common Upper Extremity Conditions.* J Am Acad Orthop Surg, 2019. **27**(15): p. 575-580.

12. Valentine, K.D., et al., *Assessing the quality of shared decision making for elective orthopedic surgery across a large healthcare system: cross-sectional survey study.* BMC Musculoskeletal Disorders, 2021. **22**(1): p. 967.

13. Evans, R.S., *Electronic Health Records: Then, Now, and in the Future.* Yearb Med

Inform, 2016. **Suppl 1**(Suppl 1): p. S48-61.

14. Micky, T., *"EHR Evolution: Policy and Legislation Forces Changing the EHR".* Journal of AHIMA, October 2012. **83**(no 10): p. 24 - 29.

15. Technology., O.o.t.N.C.f.H.I., *Adoption of Electronic Health Records by Hospital Service Type 2019-2021*, in *Health IT Quick Stat #60*. April 2022: www.healthit.gov.

16. Raghupathi, W. and V. Raghupathi, *Big data analytics in healthcare: promise and potential.* Health Inf Sci Syst, 2014. **2**: p. 3.

17. Assale, M., et al., *The Revival of the Notes Field: Leveraging the Unstructured Content in Electronic Health Records.* Front Med (Lausanne), 2019. **6**: p. 66.

18. Hanson, J.L., et al., *Quality of outpatient clinical notes: a stakeholder definition derived through qualitative research.* BMC Health Serv Res, 2012. **12**: p. 407.

19. Shah, R.F., S. Bini, and T. Vail, *Data for registry and quality review can be retrospectively collected using natural language processing from unstructured charts of arthroplasty patients.* Bone Joint J, 2020. **102-b**(7_Supple_B): p. 99-104.

20. Hashir, M. and R. Sawhney, *Towards unstructured mortality prediction with free-text clinical notes.* Journal of Biomedical Informatics, 2020. **108**: p. 103489.

21. Podder V, L.V., Ghassemzadeh S, *SOAP Notes.* 2021, Treasure Island (FL): StatPearls Publishing.

22. Hunter, K.M., *Doctors' Stories: The Narrative Structure of Medical Knowledge*. 1993, Princeton, N.J.: Princeton Univ. Press.

23. Resnik, P., et al., *Communication of clinically relevant information in electronic health records: a comparison between structured data and unrestricted physician language.* Perspectives in Health Information Management, 2008. **2008**: p. 1-10.

24. Boag, W., et al., *What's in a note? unpacking predictive value in clinical note representations.* AMIA Summits on Translational Science Proceedings, 2018. **2018**: p. 26.

25. Varpio, L., et al., *The EHR and building the patient's story: A qualitative investigation of how EHR use obstructs a vital clinical activity.* International Journal of Medical Informatics, 2015. **84**(12): p. 1019-1028.

26. Vest, J.R., et al., *Using structured and unstructured data to identify patients' need for services that address the social determinants of health.* Int J Med Inform, 2017. **107**: p. 101-106.

27. Gagnier, J.J., H. Morgenstern, and P. Kellam, *A retrospective cohort study of adverse events in patients undergoing orthopaedic surgery.* Patient Safety in Surgery, 2017. **11**(1):

p. 15.

28. Obeid, J., et al., *Automated detection of altered mental status in emergency department clinical notes: a deep learning approach.* BMC Med Inform Decis Mak, 2019. **19**(1): p. 164.

29. Kreimeyer, K., et al., *Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review.* Journal of Biomedical Informatics, 2017. **73**: p. 14-29.

30. Wang, Y., et al., *Clinical information extraction applications: A literature review.* Journal of Biomedical Informatics, 2018. **77**: p. 34-49.

31. Liu, P., et al., *Application of artificial intelligence technology in the field of orthopedics: a narrative review.* Artificial Intelligence Review, 2024. **57**(1): p. 13.

32. Velupillai, S., et al., *Using clinical Natural Language Processing for health outcomes research: Overview and actionable suggestions for future advances.* J Biomed Inform, 2018. **88**: p. 11-19.

33. Choi, Y.J. and H.J. Ra, *Patient Satisfaction after Total Knee Arthroplasty.* Knee Surg Relat Res, 2016. **28**(1): p. 1-15.

34. Ayers, D. and K. Bozic, *The importance of outcome measurement in orthopaedics.* Clinical Orthopaedics and Related Research, 2013. **471**(11): p. 3409-3411.

35. Khatib, Y., et al., *Patient satisfaction and perception of success after total knee arthroplasty are more strongly associated with patient factors and complications than surgical or anaesthetic factors.* Knee Surgery, Sports Traumatology, Arthroscopy, 2020. **28**(10): p. 3156-3163.

36. Gupta, R., et al., *Surgeon-specific factors have a larger impact on decision-making for the management of proximal humerus fractures than patient-specific factors: a prospective cohort study.* J Shoulder Elbow Surg, 2023. **32**(8): p. 1701-1709.

37. Patel, A.H., et al., *Trending a decade of proximal humerus fracture management in older adults.* JSES International, 2022. **6**(1): p. 137-143.

38. Davey, M.S., et al., *Management options for proximal humerus fractures - A systematic review & network meta-analysis of randomized control trials.* Injury, 2022. **53**(2): p. 244-249.

39. Spross, C., et al., *Evidence-based algorithm to treat patients with proximal humerus fractures-a prospective study with early clinical and overall performance results.* J Shoulder Elbow Surg, 2019. **28**(6): p. 1022-1032.

40. Misra, S., et al., *Practice guidelines for proximal humeral fractures.* J Clin Orthop Trauma, 2019. **10**(3): p. 631-633.

41. Schumaier, A. and B. Grawe, *Proximal Humerus Fractures: Evaluation and Management in the Elderly Patient.* Geriatr Orthop Surg Rehabil, 2018. **9**: p. 2151458517750516.

42. LaMartina, J., 2nd, et al., *Difficulty in decision making in the treatment of displaced proximal humerus fractures: the effect of uncertainty on surgical outcomes.* J Shoulder Elbow Surg, 2018. **27**(3): p. 470-477.

43. Sabharwal, S., et al., *Trials based on specific fracture configuration and surgical procedures likely to be more relevant for decision making in the management of fractures of the proximal humerus: Findings of a meta-analysis.* Bone Joint Res, 2016. **5**(10): p. 470-480.

44. Rangan, A., et al., *Surgical vs nonsurgical treatment of adults with displaced fractures of the proximal humerus: the PROFHER randomized clinical trial.* Jama, 2015. **313**(10): p. 1037-47.

45. Baker, H.P., et al., *Management of Proximal Humerus Fractures in Adults—A Scoping Review.* Journal of Clinical Medicine, 2022. **11**(20): p. 6140.

46. Wright, A.A., et al., *Defining patient acceptable symptom state thresholds for commonly used patient reported outcomes measures in general orthopedic practice.* Man Ther, 2015. **20**(6): p. 814-9.

47. Katz, N.P., F.C. Paillard, and E. Ekman, *Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions.* Journal of Orthopaedic Surgery and Research, 2015. **10**(1): p. 24.

48. Harris, P.A., et al., *The REDCap consortium: Building an international community of software platform partners.* Journal of Biomedical Informatics, 2019. **95**: p. 103208.

49. Harris, P.A., et al., *Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support.* J Biomed Inform, 2009. **42**(2): p. 377-81.

50. McHugh, M.L., *Interrater reliability: the kappa statistic.* . Biochemia medica, 2012. **22**: p. 276–282.

51. Bangdiwala, S.I. and V. Shankar, *The agreement chart.* BMC Medical Research Methodology, 2013. **13**(1): p. 97.

52. Fleiss, J.L., *Measuring nominal scale agreement among many raters.* Psychological bulletin, 1971. **76**(5): p. 378-382.

53. Grove, A., A. Clarke, and G. Currie, *How are evidence and knowledge used in orthopaedic decision-making? Three comparative case studies of different approaches to implementation of clinical guidance in practice.* Implementation Science, 2018. **13**(1): p. 75.

54. Cillessen, F., et al., *Factors affecting the manual linking of clinical progress notes to problems in daily clinical practice: A retrospective quantitative analysis and cross sectional survey.* Health Informatics Journal, 2021. **27**(2): p. 14604582211007534.

55. Röttele, N., et al., *Agreement of physician and patient ratings of communication in medical encounters: A systematic review and meta-analysis of interrater agreement.* Patient Education and Counseling, 2020. **103**(10): p. 1873-1882.

56. Walker, J., et al., *OpenNotes After 7 Years: Patient Experiences With Ongoing Access to Their Clinicians' Outpatient Visit Notes.* J Med Internet Res, 2019. **21**(5): p. e13876.

57. Gotlieb, R., et al., *Accuracy in Patient Understanding of Common Medical Phrases.* JAMA Netw Open, 2022. **5**(11): p. e2242972.

58. Rosenbloom, S.T., et al., *Data from clinical notes: a perspective on the tension between structure and flexible documentation.* Journal of the American Medical Informatics Association, 2011. **18**(2): p. 181-186.

59. Boag, W., et al., *What's in a Note? Unpacking Predictive Value in Clinical Note Representations.* AMIA Jt Summits Transl Sci Proc, 2018. **2017**: p. 26-34.

60. Song, Z., et al., *Physician Practice Pattern Variations in Common Clinical Scenarios Within 5 US Metropolitan Areas.* JAMA Health Forum, 2022. **3**(1): p. e214698.

61. Ayers, D.C. and K.J. Bozic, *The importance of outcome measurement in orthopaedics.* Clin Orthop Relat Res, 2013. **471**(11): p. 3409-11.

62. Harreld, K., et al., *Correlation of subjective and objective measures before and after shoulder arthroplasty.* Orthopedics, 2013. **36**(6): p. 808-14.

63. Bozkurt, S., et al., *Phenotyping severity of patient-centered outcomes using clinical notes: A prostate cancer use case.* Learning Health Systems, 2020. **4**(4): p. e10237.

64. Košir, U., et al., *Patient-physician agreement on function and pain is associated with long-term outcomes in sarcoma: findings from a longitudinal study.* J Cancer Surviv, 2023.

65. Swarup, I., et al., *Patient expectations and satisfaction in orthopaedic surgery: A review of the literature.* J Clin Orthop Trauma, 2019. **10**(4): p. 755-760.

66. Schroeder, H.S., et al., *Perception of Goals and Expected Outcomes in Older Hip Fracture Patients and Their Medical Staff: A Cross Sectional Study.* Geriatr Orthop Surg Rehabil, 2023. **14**: p. 21514593231202735.

67. Archer, K.R., et al., *Factors associated with surgeon referral for physical therapy in patients with traumatic lower-extremity injury: results of a national survey of orthopedic trauma surgeons.* Phys Ther, 2009. **89**(9): p. 893-905.

68. Karpinski, K., et al., *Differences in Patients' and Surgeons' Expectations before Shoulder Stabilization Surgery.* J Clin Med, 2021. **10**(20).

69. Schwartz, C.E., A. Ayandeh, and J.A. Finkelstein, *When patients and surgeons disagree about surgical outcome: investigating patient factors and chart note communication.* Health and Quality of Life Outcomes, 2015. **13**(1): p. 161.

70. Pattabhiraman, N.K., S.C. Barnes, and H.P. Singh, *Agreement between patient-based and clinician-based assessment of the shoulder.* Journal of Shoulder and Elbow Surgery, 2018. **27**(3): p. e59-e67.

71. Belayneh, R., et al., *The role of patients' overall expectations of health on outcomes following proximal humerus fracture repair.* Orthop Traumatol Surg Res, 2021. **107**(8): p. 103043.

72. Whitebird, R.R., et al., *Personalized outcomes for hip and knee replacement: the patients point of view.* Journal of Patient-Reported Outcomes, 2021. **5**(1): p. 116.

73. Dy, C.J., et al., *Variations in surgeons' recovery expectations for patients undergoing total joint arthroplasty: a survey of the AAHKS membership.* J Arthroplasty, 2013. **28**(3): p. 401-5.

74. Henry, S.G., et al., *Goals of Chronic Pain Management: Do Patients and Primary Care Physicians Agree and Does it Matter?* The Clinical Journal of Pain, 2017. **33**(11).

75. Phillips, D., et al., *A Review of Orthopedic Resident Outpatient Notes Are Perceptions of the EMR Reflected in Current Documentation Practices?* Bull Hosp Jt Dis (2013), 2019. **77**(3): p. 194-199.

76. Lutfey, K.E., et al., *How are patient characteristics relevant for physicians' clinical decision making in diabetes? An analysis of qualitative results from a cross-national factorial experiment.* Soc Sci Med, 2008. **67**(9): p. 1391-9.

77. Janse, A.J., et al., *Quality of life: patients and doctors don't always agree: a meta-analysis.* Journal of Clinical Epidemiology, 2004. **57**(7): p. 653-661.

78. Rotenstein, L.S., et al., *Gender Differences in Time Spent on Documentation and the Electronic Health Record in a Large Ambulatory Network.* JAMA Network Open, 2022. **5**(3): p. e223935-e223935.

79. Park, J., et al., *Physician Use of Stigmatizing Language in Patient Medical Records.* JAMA Network Open, 2021. **4**(7): p. e2117052-e2117052.

80. Sneeuw, K.C., M.A. Sprangers, and N.K. Aaronson, *The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease.* J Clin Epidemiol, 2002. **55**(11): p. 1130-43.

81. Field, D., et al., *Terminal illness: views of patients and their lay carers.* Palliat Med,

1995. **9**(1): p. 45-54.

82. Jones, J.M., et al., *Assessing agreement between terminally ill cancer patients' reports of their quality of life and family caregiver and palliative care physician proxy ratings.* J Pain Symptom Manage, 2011. **42**(3): p. 354-65.

83. Marcano-Fernández, F., et al., *A Systematic Review of Outcome Measures in Orthopaedic Trauma Trials: What Are We Measuring?* Indian J Orthop, 2022. **56**(8): p. 1316-1326.

84. Lavallee, D.C., et al., *One report, multiple aims: orthopedic surgeons vary how they use patient-reported outcomes with patients.* Quality of Life Research, 2023. **32**(2): p. 425-433.

85. Hashir, M. and R. Sawhney, *Towards unstructured mortality prediction with free-text clinical notes.* J Biomed Inform, 2020. **108**: p. 103489.

86. Coran, J.J., T. Koropeckyj-Cox, and C.L. Arnold, *Are Physicians and Patients in Agreement? Exploring Dyadic Concordance.* Health Education & Behavior, 2013. **40**(5): p. 603-611.