Remotely-captured, free-text responses track with patient health states in chronic pain

1st Jenna M. Reinen
Digital Health, IBM Research
Yorktown Heights, NY, USA
jenna.reinen@ibm.com

2nd Carla Agurto
Digital Health, IBM Research
Yorktown Heights, NY, USA
carla.agurto@ibm.com

3rd Guillermo Cecchi
Digital Health, IBM Research
Yorktown Heights, NY, USA
gcecchi@us.ibm.com

4th Jeffrey L. Rogers
Digital Health, IBM Research
Yorktown Heights, NY, USA
jeffrogers@us.ibm.com

5th NAVITAS and ENVISION Studies Physician Author Group
Clinical Research, Boston Scientific
Valencia, CA, USA

6th Boston Scientific Research Scientists Consortium
Data Research and Engineering, Boston Scientific
Valencia, CA, USA

***Abstract*—Chronic pain (CP) is a complex, multidimensional condition characterized by physiological and psychological components. Advances in digital health have helped improve our understanding of CP by remotely tracking multiple symptoms on a more granular time scale, allowing for a more comprehensive representation of patient experience. Concurrently, natural language processing (NLP) holds great promise in its ability to assess cognition, emotion, and acoustic properties to identify symptoms and disease. Here, we highlight the ability to track daily health and wellness in CP patients (n=206, over 15k samples) using sentiment features derived from free-text messages via smartphone applications. Using NLP, we quantified positive and negative sentiment and emotion in unstructured text messages using Sentence Transformers and Watson Natural Language Understanding. In both tools, sentiment and emotion track with a summary metric, termed as Pain Patient States, which have demonstrated prior success in describing a patient's overall well-being. This helps validate language analysis as a snapshot representation of health, and indicates that Pain Patient States may track with emotion in CP.**

***Index Terms*—chronic pain, natural language processing, sentiment analysis, sentence transformers**

I. Introduction

Chronic pain (CP) is not a simple sensory experience, but is modulated by physiological, cognitive, psychological, and societal factors [1]. Historically, it has been difficult to monitor and integrate many symptoms together, so pain has been assessed on a rudimentary, 10-point scale. Technological advances in digital health and AI have helped to address this issue. First, device technology including sensor and smartphone apps can now capture more data with greater temporal granularity and with less patient burden. Second, machine learning algorithms can identify patterns in complex data with greater ease relative to traditional statistical approaches. For instance, recent findings have shown that using a computational approach, multiple symptoms in CP patients participating in clinical trials may be characterized by a single, interpretable metric representing a holistic measure of health, and linked to critical function-related outcomes [2]. These *Pain Patient States* (here "pain states") were used in clinical trials to track health progression over time, and as an outcome variable in a recommendations engine aimed to improve care [3].

Still, there is room to improve the ability to capture feedback from patients in a way that is quick and minimally invasive. Here, we aimed to develop a tool by which pain patients in a clinical trial could easily provide feedback about their health. Using a smart phone application, we implemented a data capture system that administers question prompts to patients about their symptoms, health, recommendations received, treatment experience, and general wellness [4]. We leveraged Natural Language Processing/Understanding (NLP/NLU) analytics to derive sentiment from these responses, both as binary and as continuous descriptors, and compared them to metrics derived from Pain Patient States [2]. We will discuss the advantages of using language to assess health in CP, and future directions in computational linguistics for digital health.

II. Methods

A. Study, Participants, and Data Collection

Participants were recruited for longitudinal, multi-center clinical studies in US pain clinics treating chronic lower back and leg pain using Spinal Cord Stimulation (ENVISION Study, Clinicaltrials.gov ID: NCT03240588). Data were comprised of questionnaire responses administered and collected via a smartphone application through a custom-designed digital health ecosystem (Boston Scientific, Valencia, CA) This included pain, mood, sleep, alertness, medication use, and activity. Participants were asked to wear a smartwatch to assess mobility using actigraphy data (Galaxy Watch S2, Samsung USA, Menlo Park, CA with custom application from Boston Scientific). Additional details are reported elsewhere [2].

B. Pain States Analysis

The model used to assign pain states to a given set of patient data was developed in prior analyses, and a detailed description is found elsewhere [2]. Briefly, questionnaire and actigraphy data were downloaded, scaled, and normalized across each day and individual. A k-means clustering algorithm was

979-8-3503-6857-4/24/$31.00 ©2024 IEEE
DOI 10.1109/ICDH62654.2024.00037

implemented to examine solutions for up to k=10. Standard stability analyses were used to choose the optimal k. To validate the model and rank the clusters, the clustering results were visually inspected and the cluster centroid distances were compared to disability and quality of life scores.

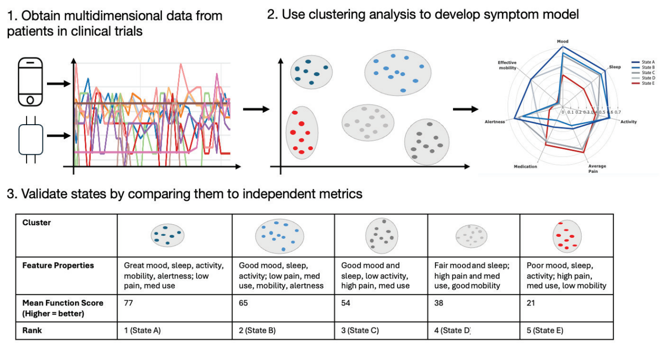

Fig. 1. Pain states were based on a model calculated in prior analyses. 1. Smartphone-based questionnaire and actigraphy data were collected from CP patients in a clinical trial. 2. Clustering analyses were used to examine symptom groupings. 3. Clusters feature means were examined and compared to independent metrics collected at the same time of the states assignments to label and rank the clusters. Here, hypothetical mean values for a functional score illustrates the method used to rank states from best (A) to worst (E).

C. Natural Language Processing and Free-Text Analysis

Free-text data were collected via the smartphone application in response to question prompts [4]. Data were cleaned for quality and availability. Next, we applied two sentiment feature extraction tools to ensure convergence across methods. First, we obtained a binary label of sentiment based on the default sentiment analysis using transformers from the Hugging Face Hub [5] by extracting the number of positive or negative labels per state. We then matched the state assignment on a given day to the free-text data for that same day for each participant. A number of occurrence counts for the "positive" or "negative" overall sentiment labels was calculated for each state.

Second, we extracted specific, continuous metrics pertaining to individual emotions (sadness, joy, fear, disgust, and anger) using a cloud-based NLU tool that leverages deep learning to obtain metadata from aspects of text-based language, including entities, key words, categories, emotional sentiment, relationships, and syntax using Watson Natural Language Understanding (WNLU) [6]. Here, we examined the Pearson correlation between each cluster's centroid distance for a given day and its corresponding sentiment using the overall WNLU component score, an emotion summary where positive emotion was denoted by larger values. We repeated this analysis for the individual intensity values for sadness, fear, disgust, anger, and joy. For comparison, we also calculated a correlation with all states in which A=5 and E=1, as well as with the participant's average pain score for that day.

III. Results

A. Participant Population

The entire data sample included in the free-text analysis included 34,447 samples from 242 participants collected between November 2019 and November 2023. Once the data were cleaned to exclude missing data, the sample for the binary sentiment analysis included 15,352 samples from 206 participants. For the emotion features analysis, the sample included 14,293 samples from 200 participants.

B. Pain States Model

A detailed summary of the model (Fig. 1) used here to assign a state to novel data is described elsewhere [2]. In this work, a solution of k=5 was determined to be optimal. The best state was characterized by good health features (low pain, low medication use, good mood, etc.), and the opposite was true for the worst state. The three intermediate states showed a combination of feature levels, but were notably similar in pain level. However, by examining the relationship between cluster centroid distances and two independent outcomes (disability and quality of life/QoL), we were able to rank the states on a best-to-worst axis, where longer centroid distances for the best state (A) associated with greater disability and QoL scores, indicating that greater distance from, or dissimilarity with, the best state associated with worse outcomes. In contrast, shorter distances or greater similarity with the worst state associated with poor outcomes. Based on this, we labeled the States A-E, where A was best and E was worst.

C. Binary Sentiment Analysis

The binary sentiment labels in the data revealed that more negative values were present than positive values (19,701 negative labels vs. 14,746 positive labels in the raw data, and 8,744 negative labels and 6,608 positive labels in the cleaned data). Despite this, the number of occurrences for negative sentiment across the best to worst states A-E was consistent with the model validation and our expectations, in that the number of occurrences and percentage occurrences was smallest for the best states (A, B) and largest in the worst states (D, E). We observed the opposite effect for the positive sentiment labels, in that State A had the most positive sentiment labels and State E had the fewest (Table 1).

TABLE I
Binary Sentiment Labels Across States

State	Negative	Positive	Total Samples / State
State A	2104 (48.3%)	2249 (51.7%)	4353
State B	1815 (47.2%)	2029 (52.8%)	3844
State C	1957 (61.7%)	1214 (38.3%)	3171
State D	1889 (68.4%)	874 (31.6%)	2763
State E	979 (80.2%)	242 (19.8%)	1221

D. Watson Natural Language Understanding Emotion and Sentiment Analysis

The WNLU analysis, including both the overall and individual emotion variables, indicated that sadness and joy demonstrated a relationship to states and to pain, in the expected directions (Table 2). All of the correlations were significant and would survive correction for multiple comparisons, except for disgust, which showed a negligible relationship to states. The worst state (E) showed an association between shorter (more similar) centroid distances and higher sadness intensity values, and an association between greater (less similar) centroid distances and higher values of joy in the free-text responses. The opposite was true of the best state (A), for which short distances (greater similarity) were associated with higher ratings of joy and lower ratings of sadness. Joy and the overall WNLU score showed a similar relationship, which was expected given that higher WNLU scores reflect positive sentiment. Anger and fear showed a significant relationship to the states, but it was of weaker magnitude than sadness.

TABLE II
WATSON NATURAL LANGUAGE UNDERSTANDING FEATURES BY STATE

	States (Numeric)	State A	State B	State C	State D	State E	Pain Rating
Sadness***	0.2	0.19	0.13	0.06	-0.09	-0.16	0.13
Fear **	0.13	0.09	0.08	-0.04	-0.1	-0.11	0.12
Digust (*ns*)	0.01	-0.02	-0.02	-0.03	-0.05	-0.04	0
Anger**	0.09	0.07	0.05	-0.01	-0.07	-0.09	0.05
Joy***	-0.22	-0.2	-0.16	0.08	0.15	0.19	-0.16
WNLU***	-0.23	-0.22	-0.17	0.05	0.12	0.18	-0.16

*** all ps< 0.00001; ** all ps < 0.0001; ns = not significant; ps would not survive correction

IV. DISCUSSION

Using two NLP tools in CP clinical trial data, we examined the relationship between sentiment in free-text responses and holistic health measures (pain states). We found relationships between 1) binary sentiment and pain states, where better states associated with more positive sentiment labels, and worse states associated with more negative sentiment labels; and 2) a relationship between specific emotions and pain states, where joy associated with the best state and sadness with the worst state. These findings serve as converging evidence that the pain states reflect overall well-being, and that even short language samples analyzed by NLP can represent health. This highlights the potential of digital health tools coupled with AI to provide meaningful clinical insights that are easy, fast, and accessible.

Several aspects should be considered. In the first analysis, more ”negative” than ”positive” labels were present in the free-text responses. This was expected, given that CP likely involves a degree of distress. Future studies could examine how this ratio changes with successful treatment. Additionally, the Pearson correlation values from the second (emotion-specific) analysis were significant, but not necessarily large in magnitude, though State A r-values were notably larger than pain alone (Table 2). While correlations in behavioral science can be small but meaningful, we speculate that this is a result of our choice to maximize data by including as many free-text responses as we could, thus including non-emotional statements (e.g., technical questions, a neutral activity, one-word responses). Future analyses should explore how varying types of text messages influence sentiment scores. Despite this, the patterns in the second analysis were consistent with our expectations and the first analysis. Moreover, the findings showed a degree of specificity in that disgust was not meaningfully correlated with the text messages, an emotion we would not expect to correlate with CP.

Overall, these findings serve as evidence that remotely-captured language samples hold promise in their ability to offer brief but meaningful representations of patient experience without scheduling a clinic visit or phone call. Currently in pain management, multiple, time-consuming questionnaires and clinical assessments are used to determine an individual's wellness. We show here that NLP can reflect information directly from the patient's chosen words to inform a clinician about their status, reducing burden to the patient and clinician.

REFERENCES

[1] R. J. Gatchel, Y. B. Peng, M. L. Peters, P. N. Fuchs, and D. C. Turk, “The Biopsychosocial Approach to Chronic Pain: Scientific Advances and Future Directions,” *Psychological Bulletin*, vol. 133, no. 4, pp. 581–624, 2007.

[2] J. M. Reinen, C. Agurto, G. Cecchi, J. L. Rogers, Navitas, Envision, and B. S. R. S. Consortium, “Definition and clinical validation of Pain Patient States from high-dimensional mobile data: application to a chronic pain cohort,” in *2022 IEEE International Conference on Digital Health (ICDH)*, Jul. 2022, pp. 47–53.

[3] T. Tchrakian, M. Zayats, A. Pascale, D. Huynh, P. Parida, C. A. Rios, S. Zhuk, and J. L. Rogers, “A Recommender for the Management of Chronic Pain in Patients Undergoing Spinal Cord Stimulation,” in *2023 IEEE International Conference on Digital Health (ICDH)*, Jul. 2023, pp. 109–114. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10224746

[4] G. Cecchi, J. Reinen, C. Agurto, E. Bilal, R. Norel, J. Rogers, D. Huynh, K. Lechleiter, R. Rauck, and E. Loudermilk, “Using AI to Determine the Meaning From a Patient's Words: Quantifying Free-Text Responses,” *Neuromodulation*, vol. 26, no. 4, pp. S201–S202, 2023, publisher: Elsevier. [Online]. Available: https://www.neuromodulationjournal.org/article/S1094-7159(23)00511-1/abstract

[5] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, and A. M. Rush, “HuggingFace's Transformers: State-of-the-art Natural Language Processing,” Jul. 2020, arXiv:1910.03771 [cs]. [Online]. Available: http://arxiv.org/abs/1910.03771

[6] R. Sarikaya, G. E. Hinton, and A. Deoras, “Application of Deep Belief Networks for Natural Language Understanding,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 4, pp. 778–784, Apr. 2014, conference Name: IEEE/ACM Transactions on Audio, Speech, and Language Processing. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6737243

ACKNOWLEDGMENTS

The NAVITAS and ENVISION Studies Physician Author Group includes Richard Rauck (The Center for Clinical Research), Eric Loudermilk (PCPMG Clinical Research Unit), Julio Paez (South Lake Pain Institute), Louis Bojrab (Forest Health Medical Center), John Noles (River Cities Interventional Pain), Todd Turley (Hope Research Institute), Mohab Ibrahim and Amol Patwardhan (both of Banner University Medical Center), James Scowcroft (KC Pain Centers), Rene Przkora (University of Florida), Nathan Miller (Coastal Pain and Spinal Diagnostics), and Gassan Chaiban (Ochsner Clinic Foundation). The Boston Scientific Research Scientists Consortium includes Dat Huynh, Brad Hershey, Rex Woon, and Matt McDonald, all of the Data Research and Engineering division.