Exploring Chronic Pain Experiences: Leveraging Text and Audio Analysis to Infer Well-Being Metrics

1st Carla Agurto
IBM Research
Yorktown Heights, NY, USA
carla.agurto@ibm.com

2nd Michele Merler
IBM Research
Yorktown Heights, NY, USA
mimerler@us.ibm.com

3rd Jenna Reinen
IBM Research
Yorktown Heights, NY, USA
jenna.reinen@ibm.com

4th Pritish Parida
IBM Research
Yorktown Heights, NY, USA
prparida@us.ibm.com

5th Guillermo Cecchi
IBM Research
Yorktown Heights, NY, USA
gcecchi@us.ibm.com

6th Jeffrey L. Rogers
IBM Research
Yorktown Heights, NY, USA
jeffrogers@us.ibm.com

7th The NAVITAS and ENVISION Studies Physician Author Group
Boston Scientific
Valencia, CA

8th The Boston Scientific Research Scientists Consortium
Boston Scientific
Valencia, CA

***Abstract*—Current advancements in digital health offer the promise of novel insights into chronic pain patients by combining subjective data from questionnaires and objective measures that are broadly available. However, unstructured information from speech data, which captures patients expressing themselves in their own words, has not been thoroughly analyzed in this area. Recognizing this limitation, we have implemented an approach to analyze thousands of chronic pain patients' responses (text and voice) from a longitudinal spinal cord stimulation study, where patients were asked to answer different prompts about their day and the recommendations provided to them to improve their outcomes. We used a large language model and acoustic techniques to extract features and infer seven well-being metrics in a cross-validated approach, including an overall health status assessment and a disability test, achieving statistically significant correlations of up to Spearman r=0.46.**

***Index Terms*—Chronic pain, speech, NLP, well-being**

I. Introduction

Chronic pain affects 20% of the US population with a disorder that has complex effects on both the physical and mental states of sufferers [1]. While acute pain typically resolves after healing, chronic pain can last for several months and become persistent, thus severely impacting the quality of life of an individual [2], [3]. Currently, pain assessment is performed through the use of self-reporting rating scales, which often fall short when it comes to chronic pain [4]. One of the main reasons is that chronic pain is a complex, multidimensional experience [5], encompassing not only sensory aspects, but also emotional, cognitive and even social dimensions.

For example, a study [6] estimated that sleep problems have a prevalence of 88% in chronic pain patients. Therefore, analyzing chronic pain in individuals requires understanding sleep, mental state, and mobility. Often, self-reports of mood, sleep quality, and activities performed during the day can be captured [7], [8]. By assessing functional limitations and activity levels, healthcare providers can gain insights into the practical consequences of chronic pain on a patient's life.

Disability metrics also play a crucial role in assessing the impact of chronic pain on individuals' daily functioning and quality of life. One of the most commonly used tests for this purpose is the Oswestry Disability Index (ODI) [9]. The ODI total score provides a comprehensive assessment of functional impairment related to chronic pain, covering various domains such as walking, sitting, lifting, and sleeping. These are quantified on a scale of 0 to 100 (representing maximum disability). Other metrics that are also used to quantify the effect of chronic pain in an individual are those that measure overall health status. One common test is the EuroQol 5 Dimension 5 Level (EQ-5D-5L) [10]. This test is a self-reported survey that measures quality of life across five domains: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. In particular, we are interested in the EQ-5D Health Visual Analog Scale (VAS), also known as the EQ-VAS or EQ-5D Health VAS. This is a component of the EQ-5D questionnaire that captures an individual's overall perception of their health without focusing on specific aspects [11]. This component quantifies health status in a scale from 0 to 100 (best health).

Chronic pain can evoke various emotional responses, such as frustration, sadness, or anxiety. These emotions can influence speech patterns, causing tone, pitch, or cadence changes. For example, someone experiencing high levels of pain and frustration might speak more quickly or with a sharper tone.

979-8-3503-6857-4/24/$31.00 ©2024 IEEE
DOI 10.1109/ICDH62654.2024.00041

Another area that can impact speech is attention processing. For example, someone with chronic pain might struggle to maintain focus during a conversation due to pain-related distractions. This could manifest as pauses, interruptions, or difficulty staying on topic during speech. On top of emotions, speech can also reveal other relevant information about a patient, such as poor sleep, interference of pain in some activities, etc. Asking patients open-ended questions to patients about themselves and their feelings, coupled with speech analysis, offers an opportunity for important insights into their well-being.

Recently, large language models (LLM), which are trained in massive datasets containing vast amounts of text from the internet, books, articles, and other sources such as BERT (Bidirectional Encoder Representations from Transformers) [12], have been used successfully to capture and understand complex linguistic structures, including grammar, syntax, semantics, and even context-dependent nuances. These have been applied successfully to areas such as electronic health records [13], schizophrenia spectrum disorders [14] and dementia [15], among others.

In this paper, we report the extension of these tools to understand chronic pain. These methods allow open-ended questions captured through widely used technology to provide unique insights into the chronic pain experience. Specifically, we use patient feedback on therapy recommendations captured through a long-term study to gain insights into the effects of chronic pain. We analyze free text responses and voice recordings obtained during three years of the study. We extract NLP-based and acoustic features, and input those features into machine learning classifiers to infer several self-reported metrics (*e.g.*, pain, mood, sleep quality, alertness and pain interference), one disability metric using (ODI test), and one overall health metric (EQ-5D health VAS score).

II. Methods

A. Data and Protocol

Data were collected from in-clinic and at-home streams from two ongoing, multi-center, prospective Boston Scientific-sponsored SCS studies (NAVITAS and ENVISION) at up to 30 U.S. sites (Clinicaltrials.gov ID: NCT03240588). All participants were SCS candidates or being treated with SCS. Participants were asked, through a custom-designed clinical study version of a digital health ecosystem (MyStudyPartner+, formerly QLINIQL, Boston Scientific, Valencia, CA) to provide daily self-reported assessments of different aspects of their life such as overall pain intensity, sleep quality and mood, among others. Additionally, participants were asked to answer free text questions to gather information about how the participants were feeling and if the recommendations provided to lower their pain were useful. Additionally, similar questions were asked to the participants, but responses were acquired using voice recordings. For this work, we focused on three prompts. Table I displays those posed to participants by category (text and voice recordings).

TABLE I
PROMPTS ANSWERED BY THE PARTICIPANTS USING TWO METHODS: 1)TYPING THE RESPONSE (TEXT), AND 2)RECORDING THEIR VOICE WITH THE RESPONSE (VOICE)

Category	Prompt
Text[1]	- How was your day today? Can you share how the recommendations from the app may have made a difference? - How was your day (*e.g. activity, sleep, pain medications, mood, pain intensity)?* - Would you like to share any additional information about the importance to you of any of the following factors: overall pain, leg pain, low back pain, sleep time, sleep quality, mobility/daily activities, mood, pain medication(s), or other factors that might be important to you?
Voice	- Take some time to reflect and describe anything that has changed for you and anything that you think has been noticeable or note-worthy.Pain? Mood? Hobbies? Activities of daily living? Sleep? Socializing? - Would you like to talk about the X recommendations you received Y? (where X=None, 'Spinal Cord Stimulator', 'Behavioral', 'general', 'mobility'; Y='today', 'this week', 'over the last few days') - Would you like to talk about the spinal cord stimulation recommendation(s) you have seen in For You over the last 2 weeks?

[1]For text responses, we only reported the top three prompts that account for %99 of the samples.

B. Clinical Variables

To evaluate if the responses provided by the participants could characterize the effects of chronic pain in their daily lives, we used five self-reported measurements at home: mood and sleep quality using a scale from 1 to 5 (best); overall pain assessed using a scale from 0 to 10 (high pain); pain interference from 0 to 4 (worst); and alertness from 0 to 4 (high alertness). Additionally, we used the ODI total score ranging from 0 to 100 (worst) and EQ-5D health VAS varying from 0 to 100 (best). For these assessments, participants performed the test at the clinic and at home.

C. Feature Extraction

Text features were extracted by embedding all the responses by a participant using BERT transformers, specifically we used RoBERTa [16]. Then, we computed semantic similarity to pre-defined statements, which were selected based on topics affected by chronic pain participants, as shown in Table II. Specifically, we wanted to see the participant's evolution, so we asked whether there was improvement or deterioration on 11 general topics (*e.g.*, mood) and three specific topics related to the study (*e.g.*, recommendations). In addition, we computed the difference between positive and negative statements, obtaining a feature representation of 42 dimensions for each response.

To process audio responses, we extracted acoustic features using the Geneva Minimalistic Acoustic Parameter Set (GeMAPS v2.0) [17] from the OpenSMILE open-source toolbox [18]. This feature set contains 80 different features that capture properties like formant frequencies, voice stability, and mel-frequency cepstral coefficients (MFCCs) and others. Additionally, we transcribed these recordings using Whisper Open AI [19] and extracted the same text features explained at the beginning of this subsection.

D. Experimental Design

Since free text responses and self-assessment metrics (*e.g.*, ODI and EQ-5D) are acquired independently, *N.B.*, they are not necessarily performed on the same day; we associated each text response with its closest score. However, we excluded responses related to a score that were obtained more than seven days apart to ensure the temporal integrity of the questionnaire (ODI, EQ-5D) data. Once we defined this cohort, we aligned the samples with all available daily self-reported measurements of alertness, mood, overall pain, pain interference, and sleep quality. Additionally, since we were interested in evaluating whether we could infer a score based on different responses from the same subject, we averaged all samples from the same participant. Then, we standardized the features (mean = 0 and standard deviation = 1) and input them into three different regression algorithms: Lasso, Ridge, and Support Vector Regression (SVR). Parameter and feature selection were performed in the training set. In particular, we optimized alpha=[1E-2, 1E-1, 0, 1, 1E2] for Ridge and Lasso, and C =[1, 1E-1, 1E-2, 1E-3], and epsilon [.1, 1, 5, 10, 20] for SVR. To evaluate the performance of all regression models, we used a 10-fold cross-validation approach, where we stratified the folds at the subject level to avoid overfitting.

III. Results & Discussion

A. Data

From the 34450 free text responses collected in the study, we analyzed 10318 given the constraint of having available ODI and/or EQ-5D Health VAS scores. These responses correspond to 166 participants who answered one of the prompts shown in Table I from December 2019 to July 2023. More details about this cohort of participants are specified in Table III. There were 954 responses for voice recordings, but only 258 had a valid ODI and/or EQ-5D Health VAS score, corresponding to 37 subjects. Note that this cohort is a sub-cohort of the subjects in the text response cohort. When we compared the average values for each type of response, the voice response cohort had 6% more males and slightly higher pain duration. Specifically, for the variables that we inferred, we observed that overall, the average participant is somewhat skewed towards better than average health, with mood and sleep quality values being above their mean values of 3 (scale from 1 to 5), alertness being higher than 2.5 (scale from 0 to 4), and disability values being lower than 50 (scale from 0 to 100).

We also analyzed the correlations between the variables to predict as it is shown in Figure 1. We found higher correlations between ODI and pain interference (r=0.64 for text responses and r=0.70 for voice responses). This is expected given that the ODI questionnaire concentrates on how pain affects various aspects of daily living and functional abilities. We also observe that mood and sleep quality have notable correlations (r=0.58 for text responses and r=0.69 for voice responses). Similar relationships has been previously reported [20], where researchers reported that poor sleep quality could be used to predict high pain intensity the next day, and negative mood mediated this relationship. Interestingly, sleep quality and disability metrics lack association, especially for voice responses with an absolute r < 0.1.

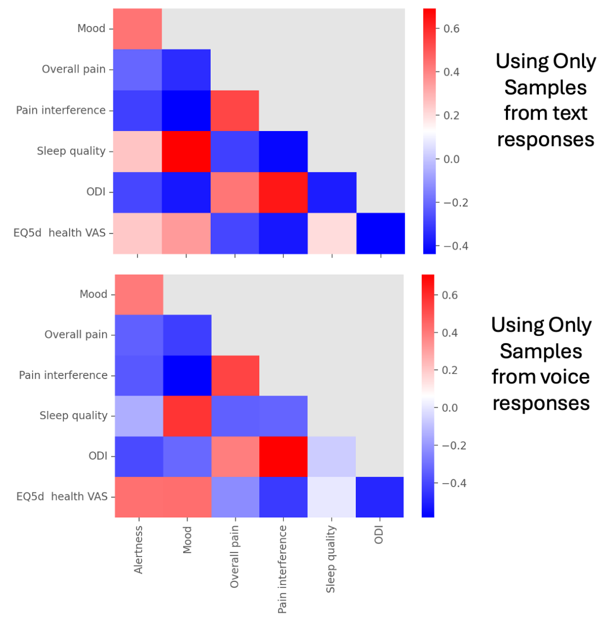

Fig. 1. Correlation between the self-reported assessment variables and tests using only text responses (above) and voice responses (below).

B. Classification experiments

Depending on the uniqueness of the participant ID and score combination, the number of samples changed in every experiment. For example, for text responses, we obtained more than 700 samples for inferring ODI and EQ-5D Health VAS while depending on the variable to infer (*e.g.*, mood) and the granularity of the assessment (*i.e.*, scales from 0 to 10), we had between 506 and 3117 samples (see Table IV). Furthermore, the number of samples was less for voice responses (range from 55 to 125 depending on the variable to infer). We also observed that higher performance were achieved for features extracted from text responses, achieving Spearman correlation coefficients of r=0.36 ($p<1E-5$) and r=0.31 ($p<1E-5$) in comparison to the ones achieved for voice recordings (r=0.29, $p<2E-2$; and r=0.29, $p<2E-2$) for ODI total and EQ-5D Health VAS scores, respectively. For other self-reported variables, high performance was achieved for mood (r=0.46, $p<1E-5$), pain interference (r=0.44, $p<1E-5$), and sleep quality (r=0.40, $p<1E-5$) for text responses; and for pain interference (r=0.45, p=6E-4) and alertness (r=0.42, p=1E-4) for voice responses. Finally, we also observed that the highest-performing regression algorithm depended on the

TABLE II

STATEMENTS USED TO COMPUTE SIMILARITY WITH THE PARTICIPANT'S RESPONSES.

Target	Sentence Positive	Sentence Negative
Pain	I have less pain intensity	I have more pain intensity
Mood	My mood is better	My mood is worse
Attentive	I am more attentive	I am less attentive
Anxiety	I am less anxious	I am more anxious
Socialize	I am socializing more	I am socializing less
Medication	I am using less medication	I am using more medication
Exercise	I am exercising more	I am exercising less
Housechores	I can do more house chores	I can do less house chores
Sleep hours	I am sleeping more	I am sleeping less
Sleep quality	I am sleeping better	I am sleeping worse
Job	It is easier to do my job	It is more difficult to do my job
Device	The device is working	The device is not working
Recommendation	The recommendation was beneficial	The recommendation was detrimental
Implementation	I did implement the recommendation	I did not implement the recommendation

TABLE III

DEMOGRAPHIC, PAIN SPECIFIC METRICS AND SELF-REPORTED VARIABLES FOR ALL PARTICIPANTS. VALUES ARE EXPRESSED AS MEAN +/- STD.

Category	Variable	Data from text responses	Data from voice responses
Demographics	Participants (Number of samples)	166 (10318)	37 (258)
	Age (years) at baseline	61.4 +/- 11.8	63.0 +/- 11
	Gender % males	34.9 %	40.5%
Pain Specific	Years since onset of pain	16.0 +/- 12.3	17.3 +/- 13.2
	Back pain	95.8 %	100%
	Unilateral lower extremity	28.9 %	35.1%
	Bilateral lower extremity	61.4 %	56.8%
	Last follow-up duration	608.2 +/- 556.3	851.6 +/- 592.2
	SCS - Yes (%)	67.5	64.9%
	Days of data used in analysis	247.9 +/- 251.6	64.5 +/- 69.7
Self-reported variables (at home)	Alertness [0-sedate 4-alert]	2.75 +/- 0.98	3.07 +/- 0.88
	Mood [1 5-best]	3.47 +/- 0.97	3.35 +/- 0.95
	Overall pain [0 10-worst]	4.44 +/- 1.99	4.91 +/- 1.19
	Pain interference [0 4-worst]	1.54 +/- 0.96	1.76 +/- 0.99
	Sleep quality [1 5-best]	3.43 +/- 1.09	3.38 +/- 1.09
Self-reported assessments (at home and clinic)	ODI total score [0 100-worst]	38.05 +/- 15.56	41.13 +/- 15.2
	EQ-5D Health VAS [0 100-best]	68.73 +/- 19.12	69.19 +/- 19.66

variable; in most cases, the Ridge regressor algorithm achieved the best performance.

Analyzing the relevant features based on normalized weights (see Figure 2), we observe that a feature correlated with high alertness is the loudness peak rate, a feature of speech tempo. Specifically, we find that high speech tempo is associated with high alertness. This finding is intuitive when considering that an alert person tends to speak faster than a sedate one, who is expected to speak more slowly. Surprisingly, top features that help infer mood are related to sleep quantity. High sleep quantity is associated with better mood. This pattern has been observed in other studies [21], suggesting that sleep quantity may be linked to the overnight recovery from affective stress, thereby fostering a positive mood in the subsequent days. Sleep quantity was also a predictor for sleep quality, but also positive socialization (*I am socializing more*). Regarding pain interference, the pain negative statement (*I have more pain intensity*) indicates higher pain interference, as expected. For this model, we also observe that features related to recommendations and device positive statements (see Table II) are also good predictors and are inversely associated to pain interference.

Additionally, we examined the weights associated with the models that provided more accurate inferences for ODI and EQ-5D Health VAS. As it was observed in Table IV, both models for voice and text using BERT features obtained performance values that were statistically significant with respect to the null hypothesis (H0: r=0) using a t-test. Figure 3 shows the top features used in the models. For ODI, we observed that the feature associated with a medication negative statement (*I am using more medication*), which appears in both types of inputs (voice and text responses), is associated with high disability. This indicates that this feature contains information that is not prompt- or response-type dependent. In the case of EQ-5D health VAS, relevant features for text and voice responses are different. For example, a medication negative statement is associated with low EQ-5D (more disability) for voice responses as we found for ODI; however for text responses, the most relevant features were derived from a job positive statement (*It is easier to do my job*). We hypothesize that individuals experiencing contentment and competence in their job roles will likely find greater satisfaction in their daily activities, consequently leading to an enhanced overall life satisfaction.

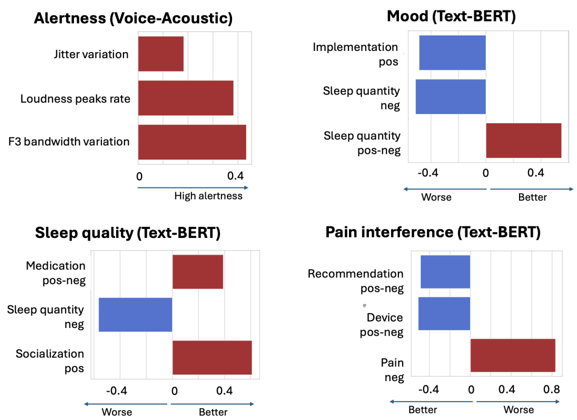

Fig. 2. Normalized weights for the models that better inferred self-reported variables. Only the first three top features according to absolute normalized weight are displayed.

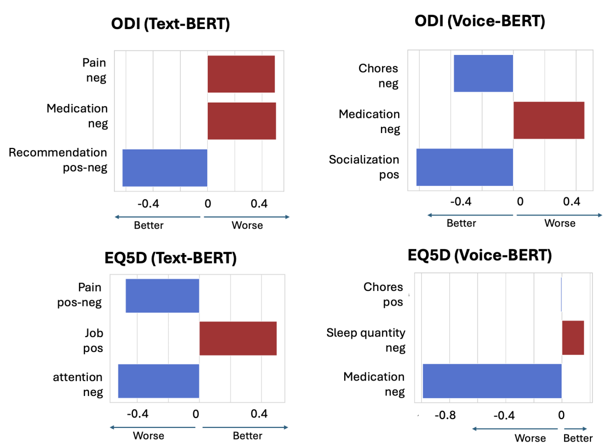

Fig. 3. Normalized weights for the best models using BERT similarity features that infer ODI total and EQ-5D Health VAS clinical variables. Only the first three top features according to absolute normalized weight are displayed.

IV. Conclusions

We find that unstructured information in chronic pain patients' experiences can be used to estimate difficult to assess aspects of their experience, including, alertness, mood, overall pain, pain interference and sleep quality with Spearman r $>$ 0.33. Importantly, we find that text or voice responses can be used to characterize ODI total and EQ-5D Health VAS. This is the first work of this size and length to determine if free speech contains relevant information of the health status of chronic pain patients.

Although we expected acoustic features not to contain information as relevant as that one embedded in content features (*i.e.*, BERT similarity), we observed that variations in speech, such as tempo, can also capture factors that are affected in chronic pain patients, such as alertness. More samples would be required to complete a more thorough analysis.

Ultimately, we believe this analysis offers a more user-friendly and streamlined option compared to conventional methods like paper-based questionnaires, which can be cumbersome and time-intensive for patients. This alternative could potentially be integrated to enhance the precision of other methods, such as those reliant on wearables. Consequently, healthcare providers could attain greater understanding into the specific challenges and limitations faced by individuals living with chronic pain, allowing for more personalized and effective treatment plans.

Acknowledgment

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 (Banner University Medical Center), Amol Patwardhan (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 (Boston Scientific, Data Research and Engineering), Kristen Lechleiter (Clinical Research, Boston Scientific), Brad Hershey (Data Research and Engineering, Boston Scientific), Rex Woon (Data Research and Engineering, Boston Scientific), and Matt McDonald (Boston Scientific, Data Research and Engineering).

References

[1] S. M. Rikard, A. E. Strahan, K. M. Schmit, and G. P. Guy, "Chronic pain among adults — united states, 2019–2021," *MMWR. Morbidity and Mortality Weekly Report*, vol. 72, no. 15, p. 379–385, Apr. 2023. [Online]. Available: http://dx.doi.org/10.15585/mmwr.mm7215a1

[2] M. A. Ashburn and P. S. Staats, "Management of chronic pain," *The Lancet*, vol. 353, no. 9167, p. 1865–1869, May 1999. [Online]. Available: http://dx.doi.org/10.1016/S0140-6736(99)04088-X

[3] M. A. Hadi, G. A. McHugh, and S. J. Closs, "Impact of chronic pain on patients' quality of life: A comparative mixed-methods study," *Journal of Patient Experience*, vol. 6, no. 2, p. 133–141, Jul. 2018. [Online]. Available: http://dx.doi.org/10.1177/2374373518786013

[4] D. A. Williams, "The importance of psychological assessment in chronic pain," *Current Opinion in Urology*, vol. 23, no. 6, p. 554–559, Nov. 2013. [Online]. Available: http://dx.doi.org/10.1097/MOU.0b013e3283652af1

[5] S. J. Love-Jones, *Pain as a Subjective, Multidimensional Experience*. Springer International Publishing, 2019, p. 141–144. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-99124-5_35

[6] M. T. Smith, M. L. Perlis, T. P. Carmody, M. S. Smith, and D. E. Giles, *Journal of Behavioral Medicine*, vol. 24, no. 1, p. 93–114, 2001. [Online]. Available: http://dx.doi.org/10.1023/a:1005690505632

[7] Y. C. Li and E. G. Hapidou, "Multidimensional visualization and analysis of chronic pain variables of patients who attended a chronic pain program," *Frontiers in Pain Research*, vol. 4, Oct. 2023. [Online]. Available: http://dx.doi.org/10.3389/fpain.2023.1125992

TABLE IV

PERFORMANCE RESULTS IN TERMS OF SPEARMAN CORRELATION COEFFICIENT (R). RESULTS ARE HIGHLIGHTED IF THE ACHIEVED R IS STATISTICALLY SIGNIFICANT WITH P-VALUE <0.05. ADDITIONALLY, TEXT IS IN BOLD FOR THE BEST PERFORMANCE MODEL PER VARIABLE TO PREDICT.

Variable to predict	Prompt Response	Type of features	Spearman (r)	p-value	Best regressor	N samples	N participants
ODI total	**Text**	**BERT similarity**	**0.36**	**<1E-5**	Lasso	**740**	**163**
	Audio	BERT similarity	0.29	2E-2	Ridge	60	32
	Audio	Acoustic	0.20	1E-1	Ridge	60	32
EQ-5D health VAS	**Text**	**BERT-similarity**	**0.31**	**<1E-5**	Lasso	**704**	**156**
	Audio	BERT similarity	0.29	2E-2	SVR	67	33
	Audio	Acoustic	0.09	5E-1	SVR	67	33
Alertness	Text	BERT-similarity	0.24	<1E-5	Lasso	893	151
	Audio	BERT similarity	0.42	2E-4	**Ridge**	78	34
	Audio	**Acoustic**	**0.42**	**1E-4**	Ridge	**78**	**34**
Mood	**Text**	**BERT-similarity**	**0.46**	**<1E-5**	**Ridge**	**818**	**166**
	Audio	BERT similarity	0.20	7E-2	SVR	84	34
	Audio	Acoustic	0.10	4E-1	SVR	84	34
Overall pain	**Text**	**BERT-similarity**	**0.33**	**<1E-5**	**Ridge**	**3117**	**164**
	Audio	BERT similarity	0.20	3E-2	SVR	125	34
	Audio	Acoustic	0.22	1E-2	SVR	125	34
Pain interference	**Text**	**BERT-similarity**	**0.44**	**<1E-5**	**Ridge**	**506**	**160**
	Audio	BERT similarity	0.28	4E-2	Ridge	55	34
	Audio	**Acoustic**	**0.45**	**6E-4**	**SVR**	**55**	**34**
Sleep quality	**Text**	**BERT-similarity**	**0.40**	<1E-5	Ridge	**614**	**164**
	Audio	BERT similarity	0.25	4E-2	SVR	68	34
	Audio	Acoustic	0.11	4E-1	SVR	68	34

[8] Z. Zambelli, E. J. Halstead, A. R. Fidalgo, and D. Dimitriou, "Good sleep quality improves the relationship between pain and depression among individuals with chronic pain," *Frontiers in Psychology*, vol. 12, May 2021. [Online]. Available: http://dx.doi.org/10.3389/fpsyg.2021.668930

[9] F. J. Roland, M., "The roland–morris disability questionnaire and the oswestry disability questionnaire." *Spine*, no. 24, pp. 3115–3124, 2000.

[10] E. Group, *Health Policy*, vol. 16, no. 3, p. 199–208, Dec. 1990. [Online]. Available: http://dx.doi.org/10.1016/0168-8510(90)90421-9

[11] R. Rabin and F. d. Charro, "Eq-sd: a measure of health status from the euroqol group," *Annals of Medicine*, vol. 33, no. 5, p. 337–343, Jan. 2001. [Online]. Available: http://dx.doi.org/10.3109/07853890109002087

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, in *Proceedings of the 2019 Conference of the North*. Association for Computational Linguistics, 2019. [Online]. Available: http://dx.doi.org/10.18653/v1/N19-1423

[13] L. Rasmy, Y. Xiang, Z. Xie, C. Tao, and D. Zhi, "Med-bert: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction," *npj Digital Medicine*, vol. 4, no. 1, May 2021. [Online]. Available: http://dx.doi.org/10.1038/s41746-021-00455-y

[14] S. X. Tang, R. Kriz, S. Cho, S. J. Park, J. Harowitz, R. E. Gur, M. T. Bhati, D. H. Wolf, J. Sedoc, and M. Y. Liberman, "Natural language processing methods are sensitive to sub-clinical linguistic differences in schizophrenia spectrum disorders," *npj Schizophrenia*, vol. 7, no. 1, May 2021. [Online]. Available: http://dx.doi.org/10.1038/s41537-021-00154-3

[15] F. Agbavor and H. Liang, "Predicting dementia from spontaneous speech using large language models," *PLOS Digital Health*, vol. 1, no. 12, p. e0000168, Dec. 2022. [Online]. Available: http://dx.doi.org/10.1371/journal.pdig.0000168

[16] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," 2019.

[17] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan, and K. P. Truong, "The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing," *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 190–202, 2016.

[18] F. Eyben, M. Wöllmer, and B. W. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor." in *ACM Multimedia*, A. D. Bimbo, S.-F. Chang, and A. W. M. Smeulders, Eds. ACM, 2010, pp. 1459–1462. [Online]. Available: http://dblp.uni-trier.de/db/conf/mm/mm2010.html#EybenWS10

[19] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," *arXiv preprint arXiv:2212.04356*, 2022.

[20] C. R. Valrie, K. M. Gil, R. Redding-Lallinger, and C. Daeschner, "Brief report: Daily mood as a mediator or moderator of the pain-sleep relationship in children with sickle cell disease," *Journal of Pediatric Psychology*, vol. 33, no. 3, p. 317–322, Sep. 2007. [Online]. Available: http://dx.doi.org/10.1093/jpepsy/jsm058

[21] A. E. Chue, K. C. Gunthert, R. W. Kim, C. A. Alfano, and A. R. Ruggiero, "The role of sleep in adolescents' daily stress recovery: Negative affect spillover and positive affect bounce-back effects," *Journal of Adolescence*, vol. 66, no. 1, p. 101–111, May 2018. [Online]. Available: http://dx.doi.org/10.1016/j.adolescence.2018.05.006