ChatGPT in the development of medical questionnaires. The example of the low back pain

Daniele Coraci (1), Maria Chiara Maccarone (1), Gianluca Regazzo (1), Giorgia Accordi (1), Jannis V. Papathanasiou (2,3), Stefano Masiero (1)

(1) Department of Neuroscience, Section of Rehabilitation, University of Padova, Padua, Italy; (2) Department of Kinesiotherapy, Faculty of Public Health, Medical University of Sofia, Sofia, Bulgaria; (3) Department of Medical Imaging, Allergology and Physiotherapy, Faculty of Dental Medicine, Medical University of Plovdiv, Plovdiv, Bulgaria.

Abstract

In the last year, Chat Generative Pre-Trained Transformer (ChatGPT), a web software based on artificial intelligence has been showing high potential in every field of knowledge. In the medical area, its possible application is an object of many studies with promising results. We performed the current study to investigate the possible usefulness of ChatGPT in assessing low back pain. We asked ChatGPT to generate a questionnaire about this clinical condition and we compared the obtained questions and results with the ones obtained by other validated questionnaires: Oswestry Disability Index, Quebec Back Pain Disability Scale, Roland-Morris Disability Questionnaire, and Numeric Rating Scale for pain. We enrolled 20 subjects with low back pain and we found important consistencies among the validated questionnaires. The ChatGPT questionnaire showed an acceptable significant correlation only with Oswestry Disability Index and Quebec Back Pain Disability Scale. ChatGPT showed some peculiarities, especially in the assessment of quality of life and medical consultation and treatments. Our study shows that ChatGPT can help evaluate patients, including multilevel perspectives. However, its power is limited, and further research and validation are required.

Key Words: ChatGPT; artificial intelligence; questionnaire; low back pain; rehabilitation; diagnosis and follow-up.

During the last year, one of the potential greatest revolutions impacting humankind has been happening: the worldwide availability of chatbots based on artificial intelligence (AI).[1] In particular, on November 30th, 2022, Chat Generative Pre-Trained Transformer (ChatGPT) was released.[2] It is an online software able to communicate in a human-similar way and to answer specific questions. Furthermore, as an AI system, it can learn from the information that the user provides and dynamically interact with the human being.[2] This type of chatbot distinguishes itself from simple web-based software because of its adaptability. Indeed, it can tailor its responses to various situations and to learn from errors and diverse inputs. This capability is advantageous for solving a wide range of problems by retrieving information from a vast dataset. Its practical applications are highly diverse and may virtually include every field of knowledge. For work activities, it can solve equations or help in computer programming; for leisure time, it can provide suggestions about a trip or invent a game; for daily life, it can support text or email writing; and so on.[3] It has, at the moment, different limitations, due to the information that the ChatGPT programmers used to train it, covering a time range until the end of 2021, and to the immaturity of some procedures. For example, ChatGPT has some difficulties in elaborating very long or highly specific texts. Additionally, too complex tasks can lead to errors and the accuracy of the responses depends on the precision of the questions.[1] Despite these current limitations, the chatbot has demonstrated remarkable potential as an assistive technology and our comprehension of its capabilities is probably only in the early stages. In the few months of its availability, even in the medical field, we have been observing a growing utilization with extraordinary outcomes. The tool can combine different information from literature and scientific databases, contributing to the interpretation of research findings or the discovery of new solutions for projects.[1]

Various studies have identified potential roles for chatbots in supporting physicians' training, particularly in summarizing patient variables and extracting relevant information from extensive datasets, like big data analysis. Moreover, for medical students, chatbots can assist in research, essay writing, or thesis work.[4] In the medical field, this AI can also support patient-physician communication, simplifying medical explanations with everyday language, assessing and consolidating symptom descriptions, and facilitating translations of medical data into various languages.[5] ChatGPT can also be beneficial for patients, providing suggestions for a health or dietary plan or indicating the most appropriate medical specialist.[6] Importantly, the chatbot underscores its restricted role in its responses, emphasizing the importance of the consultation of an irreplaceable human expert in cases that require specialized medical attention. Besides the incredible positive impact on the work and the study, the risks of misunderstood are always behind the corner. Additionally, the peril of dehumanization of medical activity and the tendency to reduce the effort toward effective work are the potential side effects of the increasing use of AI. Moreover, the utilization of this tool for certain tasks raises complex ethical concerns, like the production of scientific items sold as authors' original production but really made by a computer.[7]

Therefore, currently, ChatGPT cannot replace human doctors and scientists.[8] Nonetheless, we should avoid prejudicial judgment and we need to study all the main real effects that the chatbot may produce. Considering its medical applications, limited examples may be found in questionnaire assessments or in constructing new specific questionnaires. The existing questionnaires for understanding patients' perspectives on diseases are well-established, validated, and easy to administer. However, there is a risk of incompleteness in these questionnaires. Furthermore, as patient needs continually evolve with societal changes, updating historical questionnaires becomes desirable.

Typically, questionnaire development starts with collecting qualitative or semi-quantitative data from a large database obtained through a survey.[9] The first problems in this procedure are related to the choice of the initial questions and the answer options used for the survey, the language, and, especially, the lexicon.[10] Even the communication channel and the correct choice of the interviewed people may greatly affect the answers. After data collection, a simplification process is necessary to construct the actual questionnaire, which must undergo testing for validity and reliability.[9] Once a questionnaire passes these stages, additional processes of cross-cultural validation and adaptation are needed for translation and application in cultures and languages different from those for which the questionnaire was originally developed.[9] This brief description of the intense phases of questionnaire development suggests the possible aid of ChatGPT for this purpose. The chatbot can combine data from other existing questionnaires, and scientific and not academic information from the entire world, and can translate the document with precision.

Building upon these considerations, we aim to create a questionnaire for the assessment of [low back pain (LBP)] using ChatGPT and to evaluate its effectiveness by comparing it with routinely used validated questionnaires.

Materials and Methods

We performed an observational cross-sectional study. We enrolled subjects with a history of LBP of different severity in the last six months. Exclusion criteria were: age lower than 25 years old; recent history of trauma involving the spine; previous orthopedic, rheumatologic, and neurological diseases impacting the function of the musculoskeletal system; recent oncologic conditions with metastases; use of drugs influencing the metabolism of the bone. In order to build the questionnaire, we asked ChatGPT 3.5 to generate a questionnaire for the assessment of the LBP. The questions were formulated in Italian and the answers of the software were delivered in the same language (Supplementary file). The choice of the Italian language was determined by the necessity to test it with an Italian-speaking group of patients. ChatGPT questionnaire (ChatQ) consisted of ten questions exploring different aspects of the LBP from different points of view. In six questions the evaluation was focused on the pain perceived and its worsening factors. One question was directly aimed at the investigation of LBP impact on the quality of life (QoL). The remaining three questions explored the treatments and the medical management of LBP. Because of the different areas explored, to compare the results of ChatQ with the other tools validated in the Italian language, we consolidated the results of some questions produced by the AI to obtain a final numerical value. In particular, the single-choice and ordinal answers (for example, never, sometimes, often, always) were transformed into numerical values (in the same example, from 0 to 3). The first three questions, assessing frequency, intensity, and duration of the pain and the question about QoL were treated with this simplification. The answer about the extension of the pain was considered as a binary value (0 if the pain was located in one site, 1 if it was diffuse). The answer indicating the worsening factors was transformed in value 1 if one activity was reported and in value 0 on the other hand. The answers showing the additional symptoms were simply converted into a number indicating the count of the reported symptoms (from 0 to 4). The answers about QoL were converted into: 0 (no impact), 1 (low impact), 2 (high impact). These numbers were summed to obtain the final score of the ChatQ (possible range 0–17). The other three questions about the medical treatments were not included in the final score. The questionnaire was printed and administered to the subjects for self-compilation. The subjects were also asked to complete three validated questionnaires for back pain: Oswestry Disability Index (ODI), Quebec Back
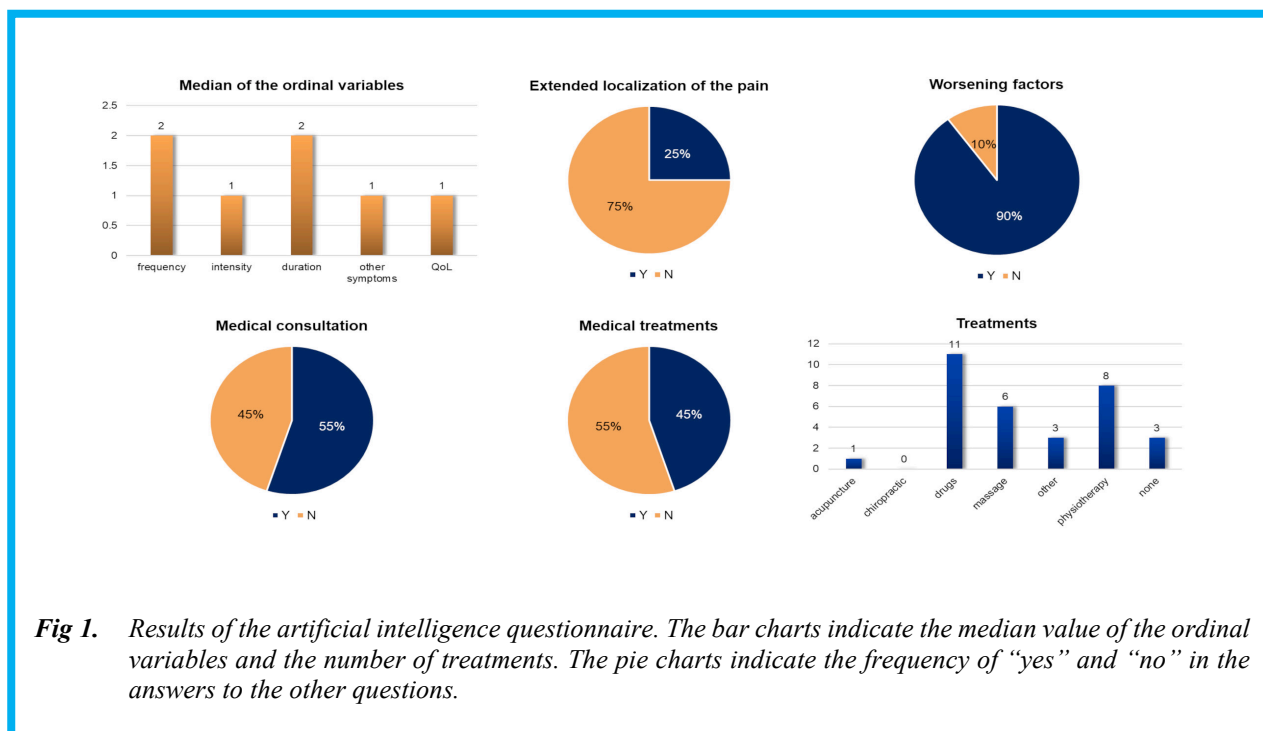
Fig 1. *Results of the artificial intelligence questionnaire. The bar charts indicate the median value of the ordinal variables and the number of treatments. The pie charts indicate the frequency of "yes" and "no" in the answers to the other questions.*

Pain Disability Scale (QBPDS), and Roland-Morris Disability Questionnaire (RMDQ). The ODI is a self-administered and Italian-language validated questionnaire designed to assess disability in individuals with LBP in both the acute and chronic phases. The test is considered the "gold standard" of low back functional outcome tools and investigates how LBP affects different aspects of daily life. Each section contains within it 6 possible answers (from 0, no disability, to 5, maximal disability). The final score is expressed as a percentage of disability (from 0% to 100%).[11] The QBPDS measures the difficulties in everyday activities for people with LBP. The questionnaire is composed of 20 items and each activity is scored on a Likert scale from 0 to 5 (0 = no difficulty and 5 = total inability). The final score can range from 0 to 100.[12] The RMDQ is designed to assess self-rated physical disability caused by LBP. The patient is instructed to put a mark next to each appropriate statement. The final score corresponds to the total number of marked statements, with a range from 0 to 24.[13] Finally, a generic evaluation of the pain, using the numerical rating scale (NRS), in which patients are asked to circle the number between 0 and 10, was applied.[14] General clinical data about gender, age, and body mass index were also collected. All data were gathered in anonymous form and the questionnaire inspector was blind to the patients' answers. This study was conducted in adherence to the principles outlined in the Declaration of Helsinki. The enrolled subjects signed an informed consent of their data treatment. The final scores of ChatQ (as described) and of the other instruments were summarized with median values. Additionally, the ordinal answers of ChatQ (frequency, intensity, duration, presence of other symptoms, impact on QoL) were summarized as median values and shown as bar charts. The binary answers (yes or no) about the extended localization of the pain, the presence of worsening factors, and the medical consultation and treatments were summarized as frequency values and pie charts. Finally, a bar chart was used to show the number of times a treatment was reported. The final scores of the various questionnaires were compared by Spearman's correlation analysis, because of the type of the data. Statistical significance was set as $p < 0.05$. The analysis was performed by the freeware software Jamovi 2.3.18. Finally, a comparison of the different main areas explored by the different questionnaires was performed, studying the most common words used in the questions and answers, by the use of TXM 0.8.0 freeware software.[15]

Results

A total of 20 subjects were enrolled (13 women, median age of 41 years, age range of 27–78 years, and median body mass index of 23.41 kg/m^2). All the patients were able to answer the administered questionnaires without assistance. The questions generated by the AI did not request specific explanations or adaptations. Median values and minimum-maximum results of the questionnaires were the following: ChatQ 8/17 (4/17–12/17); ODI 12% (0%–42%); QBPDS 9/100 (0/100–45/100); RMDQ 3/24 (0/24–11/24); NRS 4/10 (0/10–8/10). Considering the single items of the ChatQ, no subjects referred absence of LBP, in one single case the reported frequency was "always", and the median value of this item was "often". The median intensity score was classified as moderate in two cases and as severe in another. The median reported pain duration was 4-6
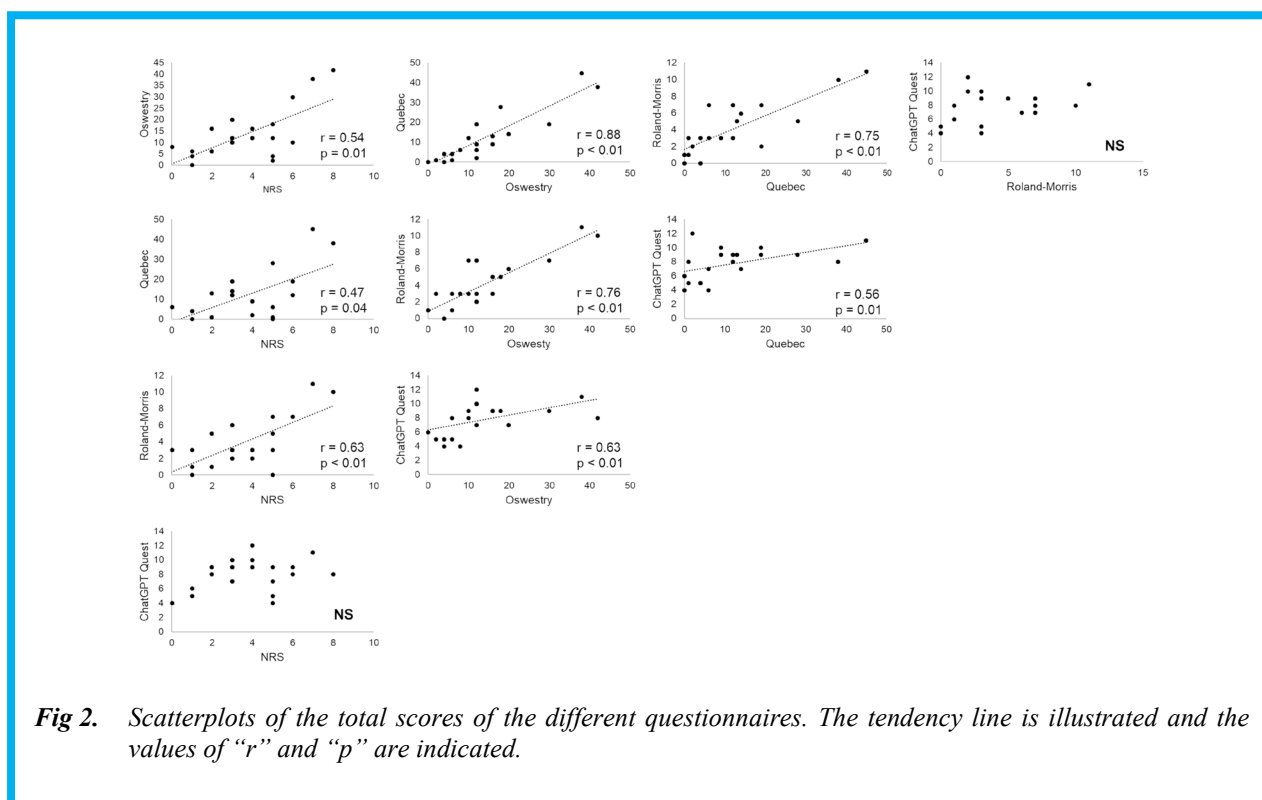
Fig 2. *Scatterplots of the total scores of the different questionnaires. The tendency line is illustrated and the values of "r" and "p" are indicated.*

hours, but for eight subjects it was more than 6 hours. Three additional symptoms (paresthesia, lower limb pain, and neck pain) were observed in a single case, compared to a median value of one, and the most frequent location being the neck. Only two patients considered their LBP severely impacted their QoL, while the low impact on QoL represented the median value. In 25% of the cases, the patients described their pain as extended in their whole trunk, and 90% of the cases reported worsening factors, mostly consisting of prolonged positions. In 55% of the cases, a medical consultation was requested; however, only 45% of the responses disclosed the medical treatments. Lastly, medication was the most commonly prescribed treatment, followed by massage and physiotherapy. Although the ChatGPT questionnaire considered chiropractic as a medical treatment, no patient admitted the use of this approach (Figure 1). The correlation analysis showed high statistical significance with strong to very strong correlation degrees among ODI, QBPDS, and RMDQ and between this latter and NRS. The correlation between NRS and ODI or QBPDS was statistically significant but of moderate level. A strong significant correlation was found between ODI and ChatQ and a moderate correlation between QBPDS and ChatQ. No statistical correlations were present between ChatQ and RMDQ or NRS. In both cases, the scatterplots showed a tendency to a consistent mutual increase of the two analyzed variables (Figure 2). The main semantic fields investigated by the various questionnaires showed common families of elements related to pain. In particular, in all the questionnaires,

including ChatQ, specific questions about actions and/or movements that can impact the pain were administered. In QBPDS, the assessment of sleep quality was added, while ODI and RMDQ additionally included questions about social relations. Only in ChatQ, specific questions about the medicalizations and the treatments were present (Figure 3).

Discussion

Our results suggest that ChatQ, specifically designed for assessing LBP, can play a potentially effective role in clinical evaluations. The results were consistent with a low level of disability in most of the subjects, while, in three subjects, the LBP was moderately severe, according to the usual validated scales. ChatQ demonstrated itself as a reliable and valuable tool for assessing LBP. Indeed, ChatQ demonstrated some good correlations with established LBP measures and allowed to explore a relatively large range of facets related to pain. Considering the relationships among the five studied tools, as expected, the four validated instruments (ODI, QBPDS, RMDQ, and NRS) showed a robust level of correlation.[16] ChatQ presented a good correlation with ODI and an acceptable one with QBPDS, but no significant results were found for RMDQ and NRS. Therefore, the precision of ChatQ as a quantitative measure is probably low. A possible explanation for the absence of correlation with NRS may be due to the different types of pain characterization. NRS is focused on the intensity, while ChatQ adds information about frequency and duration. In fact, ChatQ covers a wide
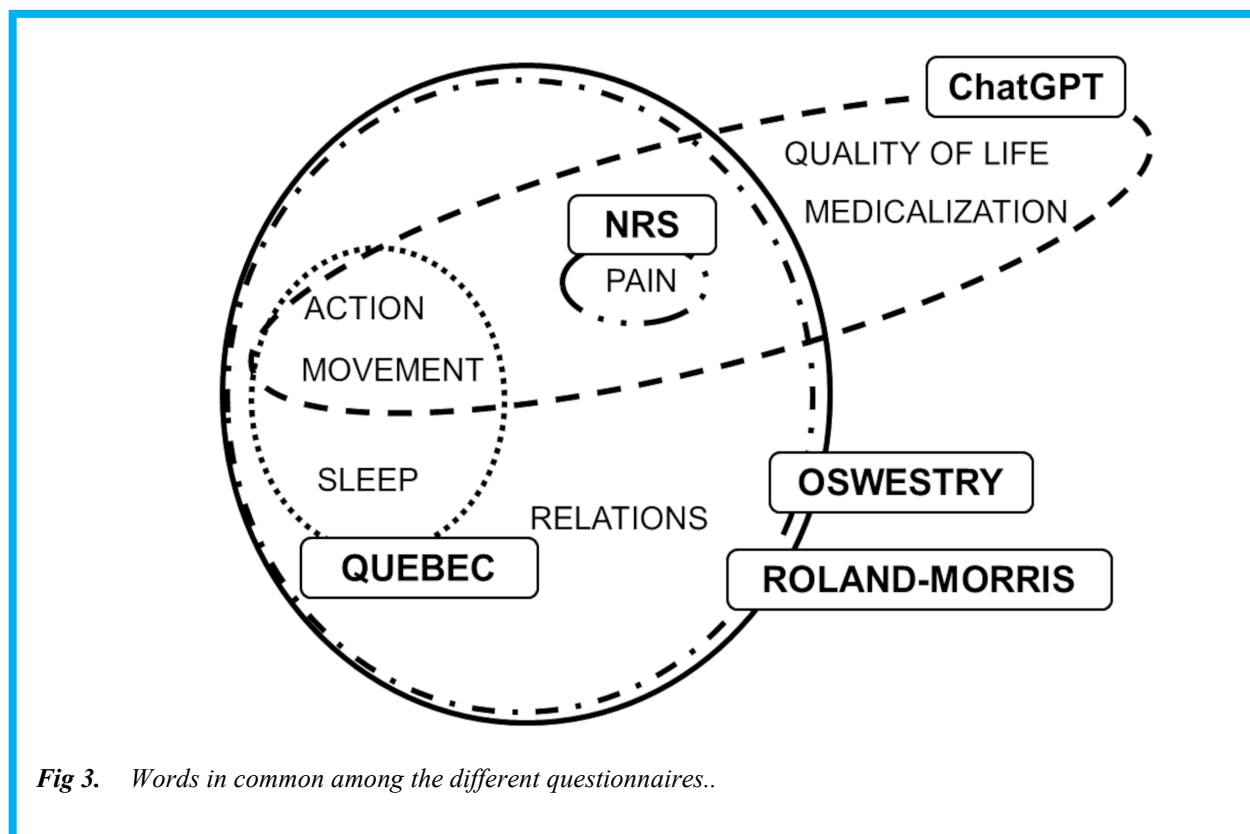
Fig 3. *Words in common among the different questionnaires..*

range of topics regarding pain perception, such as intensity, frequency, and duration. In clinical practice, the relation between these three components may be very helpful in defining the severity of LBP for both the patient and the clinician. It is possible to better define the nature and impact of pain on daily activities by focusing on all of its characteristics rather than just its intensity. Moreover, ChatQ, like other questionnaires, investigates the location of the pain and whether worsening factors are present. Different established protocols have been proposed to report pain by drawing on various body-representation templates.[17] Nevertheless, ChatQ introduces the possibility of indicating which portion of the back is affected, better defining the patient's suffering.[18] Besides the interrogations strictly focused on pain features, ChatQ explores the impact on QoL with a precise and direct question. As QoL and, in general, the impact of LBP on daily activities are two of the most important outcomes of rehabilitation, the inclusion of direct questions about QoL may help to address the evaluation.[19] However, this simple approach may not be precise enough to assess the complexity of QoL. Indeed, a well-designed questionnaire should investigate both the mental and physical aspects of life to ascertain which area is more impacted. ChatQ also queries whether there are any additional symptoms to the primary discomfort. This data could provide significant clinical information that could help define and treat the pain more precisely. Obviously, this particular information should only be used as a helpful tool in its current form. For example,

the presence of paresthesia, if reported, should guide the physician to administer further questionnaires or complete the clinical examination to depict the eventual presence of neuropathic pain.[15] ChatQ goes beyond existing questionnaires by including inquiries about medical management, which offers an innovative perspective. The three questions on this topic in ChatQ can reveal multiple facets of the patient's experience. First, the presence or not of the medical consultation may be related to the predisposition of the patient to refer to a doctor. This is a relevant point because may be a "litmus test" of the patient-physician relationship for the specific subject and may support the construction of a proficient dialogue.[20] It also provides insights into how patients perceive the severity of their pain and their health status. Furthermore, this question should be interpreted in conjunction with the inquiry about prescribed medical treatments. Indeed, medical consultation does not imply compulsory medical treatment. The absence of a prescribed medical treatment could indicate a milder condition, where a physician may adopt a wait-and-see approach or suggest minor lifestyle changes, if the back pain is not highly debilitating.[21] However, these data must be complemented by responses about the specific referred medical treatments. These answers reveal an interesting behavior of the subjects. Although a medical consultation was not reported and no medical treatments were indicated, some subjects revealed the use of self-administered therapies, in particular drugs. This information is in accordance with the vast worldwide

literature that shows the pandemical problems of auto-treatment, with the several potential side effects and the risk of misdiagnosis and difficulties, in general, of proper management.[22] Conversely, in some cases, when patients mentioned physiotherapy or massage, no medical treatment was recorded, possibly because the patients did not perceive them as medical interventions. This point should be better investigated with further and more precise surveys. Interestingly, ChatQ distinguishes between massage and physiotherapy. This may be interpreted with a high level of knowledge of the computed system. Although not totally correct, this distinction possibly underlines the differences between a basic manual therapy and a more specialized one.[23] Furthermore, the AI-generated questionnaire includes complementary and alternative medicines as possible treatment strategies.[23] Among our participants, one mentioned acupuncture, while no one referred to chiropractic care. These answers again highlight the capability of the AI to use sources from a vast database. If acupuncture has been demonstrating evidence in particular for pain control, chiropractic, at the moment, does not show a similar robustness.[24,25] However, the possibility of investigating these adjunctive traits may support the construction of new wider, and more complete questionnaires. Despite many worries about the rapidly developing new technologies, previous research has highlighted that generative AI implications in clinical practice and research are generally positively perceived among physicians in the fields of radiology, rheumatology and orthopedics.[26] Taken together, our results suggest that the questionnaire proposed by ChatGPT presents both potential advantages and limitations. Among the advantages, the evaluation of different pain variables can facilitate the identification of pain severity, while the previously validated questionnaires discriminate less clearly among these variables. AI-driven questionnaires can provide a standardized and consistent approach to LBP assessment, reducing potential variations that can occur in human-administered assessments. This standardization not only guarantees the homogeneity of criteria used to evaluate the patients, but also allows for the collection of more consistent data over time. A standardized data collection may help the rapid identification of potential issues, leading to quicker diagnosis and intervention for patients with LBP.[27] Undoubtedly, another advantage is related to the evaluation of QoL, although this should be further investigated to assess the various components of this item. Finally, the unique opportunity to investigate the relationships between the patient and medical activity deserves further detailed studies. On the other hand, the ChatQ questionnaire has significant limitations, primarily because it lacks questions about social and sexual relationships, which are crucial for assessing QoL.[28] Additionally, it does not cover sleep quality.[29] The absence of these important factors reduces the questionnaire's ability to provide a comprehensive evaluation of an individual's well-being. In the future, the inclusion of these items can improve the effectiveness of the AI-based questionnaire.

In conclusion, while still immature, ChatQ shows great potential and is quite fascinating. Despite the presence of some significant correlations with other tools, it cannot replace them. However, it can be employed to explore new subfields in the clinical evaluation of patients, due to its ability to use various types of data. Its capacity to combine data in a human-like manner may simplify the development of evaluation tools, starting with the identification of specific topics. Additionally, AI can enhance the depth of assessment processing and analysis of a wide range of data sources, such as patient-reported information, medical records, and even imaging data. Thus, ChatGPT may be used to generate, with proper input, new integrated and specific questionnaires combining the questions and the information of other tools. Therefore, while our findings indicate the promise of AI-powered questionnaires for LBP, further research and validation are necessary to ameliorate their utilization and address potential limitations. The capabilities of this type of chatbot are likely only beginning to be imagined, and further research is essential to assess its applicability for patients' benefits.

List of acronyms

AI - artificial intelligence
ChatQ - ChatGPT questionnaire
LBP - Low back pain
NRS - Numerical rating scale
ODI - Oswestry Disability Index
QBPDS - Quebec Back Pain Disability Scale
QoL - Quality of life
RMDQ - Roland-Morris Disability Questionnaire

Contributions of Authors

DC: conceptualization, data analysis, paper writing; MCM: literature analysis, paper writing; GR: data collection; GA: data collection; JVP: draft editing; SM: supervision, final approval of the manuscript.
All authors read and approved the final edited typescript.

Acknowledgments

Funding

Conflict of Interest

The authors declare they have no financial, personal, or other conflicts of interest.

Ethical Publication Statement

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Corresponding Author

Dr. Daniele Coraci, Department of Neuroscience, Rehabilitation Unit, University of Padova, Via Nicolò Giustiniani, 2, 35128 Padua, Italy.
Phone +390498216369
ORCID iD: 0000-0002-7019-9006
E-mail: daniele.coraci@unipd.it

E-mails and ORCID iD of co-authors

Maria Chiara Maccarone:
    mariachiara.maccarone@phd.unipd.it
ORCID iD: 0000-0003-2793-1334
Gianluca Regazzo: gianlucareg@gmail.com
ORCID iD: 0009-0004-6277-4199
Giorgia Accordi: giorgia.accordi@studenti.unipd.it
ORCID iD: 0009-0004-4065-3580
Jannis V. Papathanasiou: giannipap@yahoo.co.uk
ORCID iD: 0000-0003-2557-5508
Stefano Masiero: stef.masiero@unipd.it
ORCID iD: 0000-0002-0361-4898

References

1. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell. 2023 May 4;6:1169595. doi: 10.3389/frai.2023.1169595. PMID: 37215063; PMCID: PMC10192861.
2. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare (Basel). 2023 Mar 19;11(6):887. doi: 10.3390/healthcare11060887. PMID: 36981544; PMCID: PMC10048148.
3. Dergaa I, Chamari K, Zmijewski P, Ben Saad H. From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. Biol Sport. 2023 Apr;40(2):615-622. doi: 10.5114/biolsport.2023.125623. Epub 2023 Mar 15. PMID: 37077800; PMCID: PMC10108763.
4. Luo Y, Hu N. Can ChatGPT be a new educational tool in medicine? Med Clin (Barc). 2023 Oct 27;161(8):363-364. English, Spanish. doi: 10.1016/j.medcli.2023.05.018. Epub 2023 Jul 10. PMID: 37438191.
5. Gordon EB, Towbin AJ, Wingrove P, Shafique U, Haas B, Kitts AB, Feldman J, Furlan A. Enhancing patient communication with Chat-GPT in radiology: evaluating the efficacy and readability of answers to common imaging-related questions. J Am Coll Radiol. 2023 Oct 18:S1546-1440(23)00775-5. doi: 10.1016/j.jacr.2023.09.011. Epub ahead of print. PMID: 37863153.
6. Khan U. Revolutionizing Personalized Protein Energy Malnutrition Treatment: Harnessing the Power of Chat GPT. Ann Biomed Eng. 2023 Sep 20. doi: 10.1007/s10439-023-03331-w. Epub ahead of print. PMID: 37728811.
7. Ballester PL. Open Science and Software Assistance: Commentary on "Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora's Box Has Been Opened". J Med Internet Res. 2023 May 31;25:e49323. doi: 10.2196/49323. PMID: 37256656; PMCID: PMC10267777.
8. Maboloc CR. Chat GPT: the need for an ethical framework to regulate its use in education. J Public Health (Oxf). 2023 Jul 13:fdad125. doi: 10.1093/pubmed/fdad125. Epub ahead of print. PMID: 37442556.
9. Yaddanapudi S, Yaddanapudi LN. How to design a questionnaire. Indian J Anaesth. 2019 May;63(5):335-337. doi: 10.4103/ija.IJA_334_19. PMID: 31142875; PMCID: PMC6530282.
10. Coraci D, Loreti C, Fusco A, Giovannini S, Padua L. Peripheral Neuropathies Seen by Ultrasound: A Literature Analysis through Lexical Evaluation, Geographical Assessment and Graph Theory. Brain Sci. 2021 Jan 16;11(1):113. doi: 10.3390/brainsci11010113. PMID: 33467095; PMCID: PMC7829799.
11. Monticone M, Baiardi P, Ferrari S, Foti C, Mugnai R, Pillastrini P, Vanti C, Zanoli G. Development of the Italian version of the Oswestry Disability Index (ODI-I): A cross-cultural adaptation, reliability, and validity study. Spine (Phila Pa 1976). 2009 Sep 1;34(19):2090-5. doi: 10.1097/BRS.0b013e3181aa1e6b. PMID: 19730216.
12. Monticone M, Frigau L, Mola F, Rocca B, Franchignoni F, Simone Vullo S, Foti C, Chiarotto A. The Italian version of the Quebec Back Pain Disability Scale: cross-cultural adaptation, reliability and validity in patients with chronic low back pain. Eur Spine J. 2020 Mar;29(3):530-539. doi: 10.1007/s00586-019-06153-4. Epub 2019 Sep 26. PMID: 31559506.
13. Padua R, Padua L, Ceccarelli E, Romanini E, Zanoli G, Bondì R, Campi A. Italian version of the Roland Disability Questionnaire, specific for low back pain: cross-cultural adaptation and validation. Eur Spine J. 2002 Apr;11(2):126-9. doi: 10.1007/s005860100262. Epub 2001 Jul 17. PMID: 11956918; PMCID: PMC3610499.
14. Maselli F, Ciuro A, Mastrosimone R, Cannone M, Nicoli P, Signori A, Testa M. Low back pain among Italian rowers: A cross-sectional survey. J Back Musculoskelet Rehabil. 2015;28(2):365-76. doi: 10.3233/BMR-140529. PMID: 25271199.
15. Coraci D, Capobianco SV, Romano M, Calvaruso S, Vecchio M, Giovannini S, Loreti C, Fusco A, Masiero S, Santilli V, Padua L. Neuropathic Pain and Ultrasonography: A Multiperspective Literature Evaluation. Diagnostics (Basel). 2021

Sep 17;11(9):1705. doi: 10.3390/diagnostics11091 705. PMID: 34574046; PMCID: PMC8470278.

16. Garg A, Pathak H, Churyukanov MV, Uppin RB, Slobodin TM. Low back pain: critical assessment of various scales. Eur Spine J. 2020 Mar;29(3):503-518. doi: 10.1007/s00586-019-06279-5. Epub 2020 Jan 8. PMID: 31916001.

17. Coraci D, Tognolo L, Gottardello F, Posanti E, Masiero S. The use of pain drawing in medicine. A literature analysis based on graph theory. Pain Pract. 2023 Feb;23(2):220-221. doi: 10.1111/papr.13170. Epub 2022 Oct 21. PMID: 36227649.

18. Coraci D, Tognolo L, Gottardello F, Posanti E, Masiero S. The use of pain drawing in medicine. A literature analysis based on graph theory. Pain Pract. 2023 Feb;23(2):220-221. doi: 10.1111/papr.13170. Epub 2022 Oct 21. PMID: 36227649.

19. Coletti C, Acosta GF, Keslacy S, Coletti D. Exercise-mediated reinnervation of skeletal muscle in elderly people: An update. Eur J Transl Myol. 2022 Feb 28;32(1):10416. doi: 10.4081/ejtm.2022.10416. PMID: 35234025; PMCID: PMC8992679.

20. Greene J, Wolfson D. Physician Perspectives on Building Trust with Patients. Hastings Cent Rep. 2023 Sep;53 Suppl 2:S86-S90. doi: 10.1002/hast.1528.

21. Ravara B, Giuriati W, Maccarone MC, Kern H, Masiero S, Carraro U. Optimized progression of Full-Body In-Bed Gym workout: an educational case report. Eur J Transl Myol. 2023 Jun 23;33(2):11525. doi: 10.4081/ejtm.2023.11525. PMID: 37358234; PMCID: PMC10388607.

22. Abdelwahed AE, Abd-Elkader MM, Mahfouz A, Abdelmawla MO, Kabeel M, Elkot AG, Hamad MR, Ibrahim RAE, Ghallab MMI; Collaborators group; Hamza N. Prevalence and influencing factors of self-medication during the COVID-19 pandemic in the Arab region: a multinational cross-sectional study. BMC Public Health. 2023 Jan 27;23(1):180. doi: 10.1186/s12889-023-15025-y. PMID: 36707840; PMCID: PMC9880368.

23. Masiero S, Maccarone MC. Health resort therapy interventions in the COVID-19 pandemic era: what next? Int J Biometeorol. 2021 Nov;65(11):1995-1997. doi: 10.1007/s00484-021-02134-9. Epub 2021 Apr 21. PMID: 33880643; PMCID: PMC8057917.

24. Shah S, Godhardt L, Spofford C. Acupuncture and Postoperative Pain Reduction. Curr Pain Headache Rep. 2022 Jun;26(6):453-458. doi: 10.1007/s11916-022-01048-4. Epub 2022 Apr 28. PMID: 35482244.

25. Gevers-Montoro C, Provencher B, Descarreaux M, Ortega de Mues A, Piché M. Neurophysiological mechanisms of chiropractic spinal manipulation for spine pain. Eur J Pain. 2021 Aug;25(7):1429-1448. doi: 10.1002/ejp.1773. Epub 2021 Apr 15. PMID: 33786932.

26. Iyengar KP, Yousef MMA, Nune A, Sharma GK, Botchu R. Perception of Chat Generative Pre-trained Transformer (Chat-GPT) AI tool amongst MSK clinicians. J Clin Orthop Trauma. 2023 Sep 23;44:102253. doi: 10.1016/j.jcot.2023.102253. PMID: 37822477; PMCID: PMC10562840.

27. Diebo BG, Challier V, Shah NV, Kim D, Murray DP, Kelly JJ, Lafage R, Paulino CB, Passias PG, Schwab FJ, Lafage V. The Dubousset Functional Test is a Novel Assessment of Physical Function and Balance. Clin Orthop Relat Res. 2019 Oct;477(10):2307-2315. doi: 10.1097/CORR.0000 00000000820. PMID: 31135543; PMCID: PMC6999954.

28. Waxman SE, Tripp DA, Flamenbaum R. The mediating role of depression and negative partner responses in chronic low back pain and relationship satisfaction. J Pain. 2008 May;9(5):434-42. doi: 10.1016/j.jpain.2007.12.007. Epub 2008 Mar 3. PMID: 18313363.

29. Moriki K, Ogihara H, Yoshikawa K, Kikuchi K, Endo R, Sato T. Effects of sleep quality on pain, cognitive factors, central sensitization, and quality of life in patients with chronic low back pain. J Back Musculoskelet Rehabil. 2023 Aug 26. doi: 10.3233/BMR-220429. Epub ahead of print. PMID: 37694349.

Disclaimer

SUPPLEMENTARY FILE

ChatGPT questionnaire

(Italian version)

1. **Qual è la frequenza del tuo mal di schiena?**
a. **Mai**
b. **Raramente**
c. **Spesso**
d. **Sempre**

2. **Quanto è intensa la tua sensazionedi dolore al mal di schiena?**
a. **Leggero**
b. **Moderato**
c. **Grave**
d. **Molto grave**

3. **Qual è la durata media del tuo mal di schiena?**
a. **Meno di un'ora**
b. **1-3 ore**
c. **4-6 ore**
d. **Più di 6 ore**

4. **Dove si localizza il tuo mal di schiena?**
a. **Parte superiore della schiena**
b. **Parte inferiore della schiena**
c. **Schiena intera**
d. **Altro**

5. **Quali attività quotidiane aumentanoil tuo mal di schiena?**
a. **Sollevamento di pesi**
b. **Seduti o in piedi per lunghe ore**
c. **Esercizio fisico**
d. **Nessuna attività particolare**

6. **Hai mai ricevuto trattamenti medici per il tuo mal di schiena?**
a. **Sì, ho avuto cure mediche per il mio mal di schiena.**
b. **No, non ho mai avuto curemediche per il mio mal di schiena.**

7. **Quali tipi di trattamenti medici hai ricevuto per il tuo mal di schiena? (Puoi selezionare più di una risposta)**
a. **Farmaci per il dolore**
b. **Fisioterapia**
c. **Chiropratica**
d. **Agopuntura**
e. **Massaggio**
f. **Altri trattamenti**

8. **Quali altri sintomi si presentano con il tuo mal di schiena? (Puoi selezionare più di una risposta)**
a. **Dolore al collo**
b. **Dolore alle gambe**
c. **Formicolio**
d. **Nausea**
e. **Nessun altro sintomo**

9. **Il tuo mal di schiena influisce sulla tua qualità di vita?**
a. **Sì, molto**
b. **Sì, un po'**
c. **No, non mi influisce**

10. **Hai mai consultato un medico per il tuo mal di schiena?**
a. **Sì, ho consultato un medico per il mio mal di schiena.**
b. **No, non ho mai consultato unmedico per il mio mal di schiena.**

ChatGPT questionnaire (english translation)

1.   How common is your back pain?
a.   Never
b.   Rarely
c.   Often
d.   Always

2.   How intense is your back pain sensation?
a.   Light
b.   Moderate
c.   Serious
d.   Very serious

3.   How long does your back pain last on average?
a.   Less than an hour
b.   1-3 hours
c.   4-6 hours
d.   More than 6 hours

4.   Where is your back pain located?
a.   Upper back
b.   Lower back
c.   Full back
d.   Other

5.   What daily activities increase yourback pain?
a.   Weight lifting
b.   Sitting or standing for long hours
c.   Physical exercise
d.   No particular activity

6.   Have you ever received medical treatment for your back pain?
a.   Yes, I have had medical treatment for my back pain.

b.   No, I have never had medical

7. What types of treatments have you received for your back pain?

(You can select more than one answer)
a.   Pain medications
b.   Physiotherapy
c.   Chiropractic
d.   Acupuncture
e.   Massage
f.   Other treatments

8. What other symptoms come withyour back pain? (You can select more than one answer)
a.   Neck pain
b.   Leg pain
c.   Tingling
d.   Nausea
e.   No other symptoms

9.   Does your back pain affect your quality of life?
a.   Yes, a lot
b.   Yes, a bit
c.   No, it doesn't affect me

10. Have you ever consulted a doctor for your back pain?
a.   Yes, I consulted a doctorabout my back pain.
b.   No, I have never consulted adoctor for my back pain.
c.   medicaltreatment for my back pain.