# RESEARCH

# **Open Access**

# Assessing artificial intelligence-generated patient discharge information for the emergency department: a pilot study



Ruben De Rouck<sup>1,2\*</sup>, Evy Wille<sup>3</sup>, Allison Gilbert<sup>4</sup> and Nick Vermeersch<sup>1</sup>

# Abstract

**Background** Effective patient discharge information (PDI) in emergency departments (EDs) is vital and often more crucial than the diagnosis itself. Patients who are well informed at discharge tend to be more satisfied and experience better health outcomes. The combination of written and verbal instructions tends to improve patient recall. However, creating written discharge materials is both time-consuming and costly. With the emergence of generative artificial intelligence (AI) and large language models (LMMs), there is potential for the efficient production of patient discharge documents. This study aimed to investigate several predefined key performance indicators (KPIs) of Al-generated patient discharge information.

**Methods** This study focused on three significant patients' complaints in the ED: nonspecific abdominal pain, nonspecific low back pain, and fever in children. To generate the brochures, we used an English query for ChatGPT using the GPT-4 LLM and DeepL software to translate the brochures to Dutch. Five KPIs were defined to assess these PDI brochures: quality, accessibility, clarity, correctness and usability. The brochures were evaluated for each KPI by 8 experienced emergency physicians using a rating scale from 1 (very poor) to 10 (excellent). To quantify the readability of the brochures, frequently used indices were employed: the Flesch Reading Ease, Flesch-Kincaid Grade Level, Simple Measure of Gobbledygook, and Coleman-Liau Index on the translated text.

**Results** The brochures generated by ChatGPT/GPT-4 were well received, scoring an average of 7 to 8 out of 10 across all evaluated aspects. However, the results also indicated a need for some revisions to perfect these documents. Readability analysis indicated that brochures require high school- to college-level comprehension, but this is likely an overestimation due to context-specific reasons as well as features inherent to the Dutch language.

**Conclusion** Our findings indicate that AI tools such as LLM could represent a new opportunity to quickly produce patient discharge information brochures. However, human review and editing are essential to ensure accurate and reliable information. A follow-up study with more topics and validation in the intended population is necessary to assess their performance.

**Keywords** Large Language model, Patient discharge information, Artificial intelligence, Emergency department, Readability

\*Correspondence: Ruben De Rouck ruben.de.rouck@vub.be

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

# Background

Patients discharged from the emergency department (ED) need a comprehensive grasp of their eventual diagnosis, home care and subsequent follow-up. This set of instructions and education, known as patient discharge information (PDI), is crucial for successful patient care. A proper understanding of PDI is key to ensuring that patients adhere to their treatment plans. Inadequate or misunderstood PDI not only increases the risk of adverse health events and restricts patient independence but also leads to lower patient satisfaction and a greater likelihood of returning to the ED [1-3].

ED patients come from diverse backgrounds, varying in language skills, healthcare familiarity, education, and culture. Their ability to receive and process information is often compromised by health concerns, language barriers, and literacy challenges [4]. Therefore, the American National Institutes of Health (NIH) recommends that patient information should be written at a 6th grade reading level [5]. Previous research has demonstrated that patients forget approximately half of the information provided by their physicians within five minutes of their consultation, and some even report not receiving any explanation at all [6]. This issue is potentially more acute in the ED, where factors such as a chaotic environment, high staff workload, and patient literacy levels pose additional challenges [7, 8].

A systematic review by Paasche-Orlow et al. revealed that low health literacy is a widespread issue that affects patients' ability to understand information provided by healthcare professionals [9]. Despite these literacy challenges, research has demonstrated that supplementing standard care with written information can enhance memory retention, increasing recall from 47 to 58% on average [10]. A study in an urban American ED reported that the average patient reading level was at the sixthgrade level. In contrast, another study revealed that discharge instructions given to parents in a pediatric ED often require college-level reading skills [11, 12]. This disparity underscores the necessity of providing clear, comprehensive, and easily understandable written discharge instructions to improve patients' knowledge about their health conditions [13].

Adequate PDI can help to reduce ED return visits, which are associated with patient-related factors such as social problems, language problems and a lack of understanding of their diagnosis [2, 14, 15]. Several presenting complaints have been identified as contributing to ED return visits, including abdominal pain, fever and low back pain, especially when no specific diagnosis can be made [16]. Adequate PDI can help patients pay attention to certain factors that require them to return to the ED and help them overcome language and understanding issues that lead to unnecessary ED return visits. The development of PDI documents is often a timeand resource-intensive endeavor [17]. However, the increase in the use of artificial intelligence (AI) in healthcare has led to the introduction of innovative solutions. AI-driven tools, such as large language models (LLMs), are increasingly recognized for their ability to support clinical decisions, optimize workflows, and enhance patient outcomes [18].

AI has found applications in various medical fields, including diagnostic imaging, risk assessment, therapeutic drug monitoring, and patient education. AI has been employed to create visual aids alongside cardiology discharge instructions, aiming to improve patient understanding, with encouraging results [19]. An emerging use for AI is in creating patient discharge advice, a move that promises to make this crucial task more efficient [20]. Despite these advancements, the effectiveness and reliability of AI-generated PDI need thorough evaluation.

In this study, we aimed to determine the performance of an LLM in generating PDIs for nonspecific complaints frequently encountered in the ED. To our knowledge, this research is the first systematic examination of AI-generated PDI in emergency medicine, providing valuable insights for the future development and application of AI in healthcare.

## Methods

#### **PDI models**

For this study, we included three clinical scenarios to evaluate the performance of an LLM model in generating PDIs. These scenarios involved nonspecific abdominal pain, nonspecific low back pain, and pediatric (noninfant) fever. We chose to generate generic PDI brochures, assuming that the necessary and relevant diagnostic workup was prescribed and interpreted by the emergency physician, and these brochures can supplement the physician's message.

#### Nonspecific abdominal pain

Abdominal pain is a leading symptom that prompts adult visits to the ED [21]. A comprehensive approach involving medical history, clinical examination, and various imaging and laboratory tests is often needed [22]. Acute nonspecific abdominal pain (NSAP) is characterized as abdominal pain lasting less than 7 days without a definitive diagnosis following a comprehensive work-up [23]. Acute NSAP is a frequent ED diagnosis in patients with abdominal pain, with many patients being discharged under this classification. *Hoseinijad* et al. reported that 40% of patients discharged with NSAP remained undiagnosed even after one month [24]. Follow-up visits, either scheduled or unscheduled, can be crucial. *Boendemaker* et al. reported that re-evaluation within 30 h for ED patients with NSAP led to significant changes in diagnosis and treatment for approximately one-quarter of these patients [25]. In cases where no clear diagnosis is made in the ED, it is imperative to closely monitor the patient's pain and vital signs to ensure safe discharge. Providing patients with clear instructions on when to return is an essential part of this discharge procedure [26].

#### Nonspecific low back pain

Acute low back pain in the adult population is common and accounts for 2-3% of ED visits [27]. Nonspecific low back pain (NSLBP) is defined as low back pain that cannot be attributed to any specific, identifiable pathology [28]. Managing NSLBP involves balancing patient preferences with clinical evidence, typically advocating for self-management with adequate support [29]. However, evidence supporting these recommendations is limited, and patients diagnosed with NSLBP often return to the ED within 30 days, with reported rates ranging from 10 to 25%. These return visits are frequently linked to the prescription of opioid analgesics [30, 31]. Several clinical trials have shown that the effectiveness of commonly used analgesics is comparable to that of placebos [32-34]. This suggests that patient education might play a crucial role in managing return visits and the overall burden of the disease.

# Pediatric fever

Several studies have been conducted on the content of the PDI for pediatric fever, drawing from clinical experience, parental concerns, and essential medical information [35-37]. However, fever remains the most common reason for ED visits in children. The term "Fever Phobia," introduced by Schmitt in 1980, describes the persistent misconceptions and excessive fears among caregivers regarding childhood fever [38]. Decades later, this phenomenon continues to significantly contribute to ED visits for febrile children, with factors such as educational level and confusion between fever and hyperthermia playing major roles. Caregivers influenced by fever phobia are more likely to administer antipyretics unnecessarily or exceed recommended dosing intervals. Notably, over half of all parents cite healthcare professionals as their primary source of information about fever management and its implications, underscoring the critical need for effective PDI in cases of pediatric fever [39, 40]. Our study focused on children older than 90 days. This age group is chosen due to the high likelihood of hospitalization in children younger than 90 days [41].

#### LLM choice and brochure creation

LLMs are advanced AI-driven tools capable of processing and generating human-like text based on extensive training on diverse datasets. These models can understand and generate complex language constructs, making them suitable for applications in various fields, including medicine.

Currently, different LLMs are available, each designed for specific purposes. These models vary in language capability, complexity and specialization, catering to a wide range of applications from general conversation to specific professional needs. For this study, we opted for *GPT-4* (May 24th 2023 version) using the ChatGPT interface, developed by OpenAI. This model is easily available, offers a chatbot-like interface, and is trained on a large corpus of text, including medical literature [42].

The PDI brochures were created by formulating specific queries directed at the LLM, which then used the instructions to synthesize the relevant medical information into comprehensible content for patients. The query included instructions to add guidelines for over-the-counter painkiller usage and instructions for arranging follow-up care and when to return to the ED. The three PDI brochures were created in English, as most of the training data for the LLM were in English. As the study took place in a Dutch ED, the brochures were subsequently translated into Dutch using *DeepL* translation software [43]. To ensure accuracy, minor semantic errors were manually corrected, with a focus on preserving the original intent of the text. An external emergency physician, proficient in both English and Dutch and not involved in the study, validated these translations.

The three generated PDI brochures were generated after carefully crafting the specific query. The queries, outputs and machine translation of the three PDI brochures have been added as an additional file [Additional File 1.docx] to this article.

#### **Brochure evaluation**

For the assessment of PDI brochures, we enlisted a panel of 8 emergency physicians (EPs), given their primary responsibility for overseeing ED discharge. Participants in the anonymized survey were tasked with evaluating each document based on a survey using different predefined key performance indicators (KPIs): quality, accessibility, clarity, correctness of the medical information and usability. The definitions of the different KPIs are detailed in Table 1. The evaluation included a rating scale for each KPI that ranged from 1, indicating a very poor rating, to 10, signifying an excellent rating. Participants were reminded before the survey on the definitions of these KPIs. The KPIs were adapted from Rothrock et al., who performed a similar study on online information [44]. The KPIs were adapted by the investigators to the research question and were provided to the participants at the beginning of the survey after providing informed consent.

The survey was conducted from 8th July 2023 to 30th July 2023 using an online, secure, anonymized platform. The participants were selected from a single regional department comprising a team of 10 board-certified EPs with diverse expertise in emergency medicine, internal medicine, general surgery, and anesthesiology. Each physician had a minimum of 7 years of experience in emergency medicine. Because of the sample size and the risk of deanonymization, demographic data such as age, sex and individual years of ED experience were not collected.

in clinical practice.

To assess the readability and reading level, a complementary analysis was performed by means of calculating validated readability scores specifically designed for health information. We calculated 5 commonly used readability measures, namely, the Flesch Reading Ease (FRE) score, the Flesh-Kincaid grade level (FKGL), the Gunning Fog Index (GFI), the Coleman-Liau Index (CLI), and the Simple Measure of Gobbledygook (SMOG). Each analyzes text in a different manner: FRE and FKGL analyze sentence length and syllables, SMOG analyzes complex word density, GFI analyzes sentence numbers/length and complexity, and CLI analyzes characters per word and words/sentences [45]. Calculations were performed via the webFX readability tool [46].

# Ethics

Clarity

Correctness

Usability

This study was conducted in accordance with ethical standards and received approval from the Institutional Ethics Review Board of the Sint-Maria Hospital in Halle. All participating EPs were informed about the study's objectives, their rights, the confidentiality of their responses, and the voluntary nature of their participation.

# Statistical analysis

The collected ratings were statistically analyzed to determine measures of central tendency (mean and median) and dispersion (standard deviation and range) using the Python package *statsmodels version 0.12.2* [47]. The Python package *Seaborn version 0.12.1* was used to create the graphics [48].

# Results

# Participants

Of the 10 eligible emergency physicians, 8 (80.0%) participated in the survey.

#### **Brochure KPIs**

Brochure 1 (NSAP) received the highest rating for correctness, with a value of 7.8 (SD = 1.04), with quality closely behind 7.5 (SD = 1.41). Its clarity received the lowest mean score of 7.1 (SD = 1.73), and usability was rated 7.3 (SD = 0.89). The accessibility of the documents was 7.4 (SD = 1.41).

Brochure 2 (NSLBP) demonstrated the highest mean scores for accessibility, with a value of 7.9 (SD = 0.99); for clarity, with 7.8 (SD = 1.16); and for usability, with 7.8 (SD = 0.89). It had slightly lower scores for quality (7.4, SD = 0.92) and correctness (7.6, SD = 0.52).

Brochure 3 (pediatric fever) had the lowest mean score for correctness, with a value of 7.0 (SD = 1.20), and for quality, with a value of 7.1 (SD = 1.36). It scored relatively higher in clarity with 7.4 (SD = 1.69) and usability with 7.4 (SD = 1.30), with the highest rating for this document given to accessibility with 7.6 (SD = 0.92). The results are displayed in Fig. 1.

# **Readability scores**

The FRE scores suggest that all brochures are challenging to comprehend, with brochure 1 and brochure 2 falling into a very difficult category (scores of 36.8 and 36.1, respectively), while brochure 3 is marginally better at 53.9, classifying it as 'fairly difficult'. This finding is congruent with the educational levels estimated by the FKGL, where Brochure 1 and Brochure 2 require a college-level understanding (12.1 and 11.9, respectively), and Brochure 3 aligns with a 10th-grade reading level. Similarly, the Simple Measure of Gobbledygook (SMOG) index signifies a high-school-level comprehension requirement across the board, with the scores progressively decreasing from 11.2 for Brochure 1 to 9.5 for Brochure 3.

The CLI showed substantial variance among the materials, with Brochure 1 exhibiting the highest value (17.6), indicating a reading level beyond the 12th grade, while Brochure 3 presented a CLI of 12.2, which is closer to the lower secondary education level. Brochure 2 has a CLI of 15.8, suggesting a reading complexity between the two. The results for each of these measures are listed in Table 2.

How well the advice is organized and presented, and whether key instructions or

The accuracy of the medical information in

relation to the predefined clinical scenarios.

How likely it is that this advice will be used

suggestions are highlighted.

definitions attributed by the investigators		
KPI	Definition	
Quality	The overall level of comprehensiveness and accuracy of the PDI.	
Accessibility	How easily the advice can be understood by a typical patient, considering language complexity and readability.	

Table 1 Selected key performance indicators (KPIs) and their



Fig. 1 Box-and-Whisker Plots Demonstrating Evaluative Metrics of Informational Brochures Across Quality, Accessibility, Clarity, Correctness, and Usability Parameters. Brochure 1 (blue), Brochure 2 (orange), and Brochure 3 (green) are depicted

Table 2	Results	of the	readability	/ scores
---------	---------	--------	-------------	----------

Measure	Brochure 1 Nonspecific Ab- dominal Pain	Brochure 2 Nonspecific Low Back Pain	Bro- chure 3 Pediat- ric fever
FRE	36.8	36.1	53.9
FKGL	12.1	11.9	10
SMOG	11.2	10.8	9.5
CLI	17.6	15.8	12.2

FRE is Flesch Reading Ease, FKGL represents Flesh Kincaid Grade Level (FKGL), SMOG is the Simple Measure of Gobbledygook and CLI is Coleman-Liau Index

# Discussion

Our research revealed that while each document had strengths and weaknesses, they were all rated favorably, with average scores ranging between 7.0 and 7.9 across all categories. These ratings indicate that the participants generally perceived that all three documents reached an adequate level for each KPI.

However, the notably wide SDs highlighted a significant variation in how the documents were rated. This variation suggests that while there are differences in average scores among the documents, these differences might not necessarily reflect substantial disparities in the participants' perceptions of their quality, accessibility, clarity, correctness, and usability. Therefore, it is reasonable to conclude that the participants viewed the three documents as relatively comparable in terms of these key aspects.

The analysis identified two major outliers, but apart from these, there was a high level of consensus among the respondents. The specific feedback on the brochures primarily pertained to the language used and some factual inaccuracies in brochure number 3. Notably, this brochure incorrectly advised parents to start administering ibuprofen syrup to children starting at two months of age rather than the recommended safe age of three months [49]. Another point of discussion was the definition of a concerning rectal temperature: whether it should be considered from 38.0 to 38.5 °C. The LLM correctly identified 38.0 °C as the threshold for concern, aligning with the common understanding that a fever is classified as 2 standard deviations above normal body temperature [50]. However, it is important to note that temperature norms vary depending on the method of measurement, and this distinction should be clearly communicated to avoid confusion [51].

The readability assessments suggest that the language complexity of brochures necessitates a high school to college level of education for adequate comprehension. The interpretability of readability scores must be approached with caution, as their accuracy is not absolute, and scores are relative to the context and sample of the text to which they are applied. Therefore, they cannot be used interchangeably for healthcare information. However, they can be indicative and used for comparison, but their results should always be validated by other tests in the intended population [52]. The readability metrics utilized in this analysis are based on language-independent features, and empirical evidence suggests that they maintain robust correlations across languages that share structural similarities. However, peculiarities of the Dutch language, including the prevalence of lengthy words and the tendency to conjoin compounds into single entities, can lead to an underestimation of readability when assessments hinge on word length. Conversely, metrics evaluating words per sentence emerge as more predictive. Therefore, for Dutch texts, the SMOG, FRE, and FKGL metrics are more reliable indicators of readability, whereas the CLI might not provide an accurate reflection of textual accessibility [53, 54].

Across the evaluated brochures, there is a notable consistency in scores, except for brochure 3, which demonstrates superior readability across all measures. This disparity is presumed to be linked to the inherent nature of the subject matter, which likely involves the use of comparatively shorter words and less complex language constructs. When considering these assumptions, the readability indices imply that the materials are of a reading level of higher secondary education. This is still above the NIH recommendation, a finding that is consistent among studies analyzing patient education materials [55]. It should also be emphasized that readability metrics are only one facet and that evaluation of patient education materials such as the PDI should be holistic and include both quantitative and qualitative approaches [56].

#### Using LLMs in clinical medicine

The use of LLMs in healthcare shows great promise, yet concerns about ethics, transparency, and reliability remain [57]. Interestingly, despite these issues, GPT-4's language and phrasing are often seen as more empathetic than traditional medical advice from physicians [58]. Although our study was not designed for this purpose, we suspect that GPT-4 may produce content that is more

patient-centered and easier to understand than that created by emergency physicians and nurses. This was evidenced by feedback on brochure 1, where its empathetic tone was noted as feeling somewhat "out of place." The incorporation of generative AI in healthcare communications could counterbalance the often jargon-heavy and clinical language, potentially enhancing patient experience, understanding and, ultimately, outcomes.

For our study, we initially generated the PDI brochures in English and then translated them into Dutch. This approach was based on the fact that more training data are available in English than in Dutch, which we assume improves the AI model's output quality. While LLMs can create documents in various languages, we opted for AI translation tools for the conversion process, with only minor edits needed. We found the process of generating PDI brochures to be remarkably straightforward. The key lies in framing the query correctly and providing precise instructions. Once the query was finely tuned, the generation of PDI brochures was accomplished within minutes.

#### Limitations

Our experiment represented only a pilot study based on theoretical data, several important limitations warrant acknowledgment. We did not explicitly prompt the LLM to produce materials at a sixth-grade reading level, which may have contributed to a higher complexity than recommended by the NIH. Although our evaluation focused on clinically experienced emergency physicians, no patient input was obtained regarding the documents' readability and clarity. This narrow stakeholder sample limits the generalizability of our findings. Our study utilized a small sample of both participants (8 raters) and clinical scenarios (3 types of complaints), constraining the breadth of conclusions. We employed a custom five-dimension assessment framework that, while tailored to our study aims, is not a formally validated approach; as a result, direct comparisons with other LLMs or published evaluation frameworks (e.g., HELM) are challenging. Finally, although we used DeepL for translation to Dutch, further validation of whether ChatGPT's own multilingual capabilities can yield equivalent or superior translations is needed-especially considering the legal and regulatory requirements that demand human review of AIgenerated or AI-translated patient documents in some jurisdictions.

#### Future research opportunities and challenges

The quality of the PDI generated by ChatGPT using GPT-4 underscores the potential for LLMs to provide interactive, patient-centered materials. However, it is crucial to recognize that medical knowledge is neither static nor absolute, and LLMs are rapidly evolving at the

time of writing. Despite rapid advances LLMs have tendencies to produce erroneous information or "hallucinate," as observed in the febrile-child scenario, and may occasionally cite nonexistent references [59]. Once these limitations are addressed—whether through improved model training, more robust oversight, or better prompt engineering—LLMs may hold significant promise in generating PDIs finely tailored to the needs of individual patients.

Multiple avenues exist to advance this work. Future research can explicitly prompt LLMs to create PDIs specifically focused on the KPIs and at a sixth-grade reading level, thereby better enabling direct comparisons with standard AI-generated text and existing human-curated materials. Incorporating patient feedback-including individuals from diverse educational and linguistic backgrounds-will be essential to fully assess usability and refine readability. Extending both the range of clinical presentations and the reviewer pool (e.g., by involving multidisciplinary teams and patient representatives) could improve generalizability. Measuring cost-effectiveness and time savings from AI-driven generation and translation would further strengthen the evidence base. Benchmarking AI-generated brochures against humancurated counterparts would clarify relative advantages in accuracy, clarity, and time-to-production. Incorporating validated evaluation frameworks and machine translation evaluation may facilitate more standardized comparisons and allow the analysis of patient tailored LLM generated PDI [60-62].

Lastly, generation or translation of PDI information using LLMs may be subject to the legal requirement of manual overview This requirement is in place for reasons of safety, but it may mean that the time saving from using AI to generate/translate documents may not be as drastic. For instance, in the European Union, the recently adopted AI Act classifies AI systems used in medical devices as "high-risk." This classification imposes stringent requirements, including comprehensive risk assessments, quality management systems, and human oversight mechanisms, to ensure the safety and reliability of AI applications in healthcare. While the AI Act does not explicitly mandate human review of AI-translated medical documents, its emphasis on human oversight and risk management implies that such reviews are essential to maintain accuracy and compliance [63]. In the United States, the National Council on Interpreting in Health Care emphasizes compliance with Sect. 1557 of the Patient Protection and Affordable Care Act and mandates that critical medical translations undergo review by a qualified translator to ensure meaningful access for individuals with limited English proficiency [64].

#### Conclusion

In conclusion, our pilot experiment suggested the potential of LLMs, specifically ChatGPT using GPT-4, to reach adequate KPI levels for generating PDI under three conditions commonly encountered in EDs. The findings indicate that the PDIs produced were of adequate quality, suggesting that LLMs can be a valuable tool for enhancing the efficiency of creating patient education materials. This is particularly relevant in the context of emergency medicine, where time is a critical factor and the need for clear, concise, and accurate patient information is paramount.

The generated documents were generally well received by ED physicians and scored well on measures such as clarity, accessibility, and correctness. Analysis using commonly used readability measures seems to confirm these results, but the reading level of the documents is likely still above the 6th grade level recommended by the NIH. This underlines the potential of LLMs to support healthcare professionals in providing effective patient education. However, the study also revealed some challenges, most notably the instances where the AI-generated content showed potentially dangerous inaccuracies or a lack of alignment with established medical guidelines. These findings underscore the necessity for careful review and modification of AI-generated content by medical professionals before its use in real life. Although LLMs can significantly aid in the drafting of PDIs, they cannot replace the expertise and nuanced understanding of healthcare professionals.

# Abbreviations

AI	Artificial Intelligence
CLI	Coleman-Liau Index
ED	Emergency Department
EPs	Emergency Physicians
FKGL	Flesch-Kincaid Grade Level
FRE	Flesch Reading Ease
GFI	Gunning Fog Index
KPIs	Key Performance Indicators
LLMs	Large Language Models
NIH	National Institutes of Health
NSAP	Nonspecific Abdominal Pain
NSLBP	Nonspecific Low Back Pain
PDI	Patient Discharge Information
SMOG	Simple Measure of Gobbledygook

#### Supplementary Information

The online version contains supplementary material available at https://doi.or g/10.1186/s12245-025-00885-5.

Supplementary Material 1

#### Acknowledgements

The authors declare this work voided of any competing financial, professional, or personal interests from other parties.

#### Author contributions

RDR conceptualized the study and the study methodology and translated the discharge information brochures. EW checked and approved the translated

versions. Both RDR and EW wrote the manuscript. AG acted as a subject matter expert and revised the manuscript. NV helped conceptualize the study and assisted with ethics committee approval and participant recruitment. The OpenAI tool ChatGPT-4 was used to assist with phrasing and legibility of this manuscript.

#### Funding

This research was not funded. All authors are employed by their respective institutions.

#### Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

## Declarations

#### Ethics approval and consent to participate

Ethical committee approval was granted from the AZ Sint Maria hospital's ethics board. Informed consent was obtained from all individual participants included in the study.

#### **Consent for publication**

All authors have read and approved the manuscript for publication. All participants in the study consented to publication of their data before participating in the study.

#### **Competing interests**

The authors declare no competing interests.

#### Author details

<sup>1</sup>AZ Sint Maria Halle, Ziekenhuislaan 100, Halle 1500, Belgium <sup>2</sup>Research Group on Emergency and Disaster Medicine, Vrije Universiteit Brussel, Laarbeeklaan 103, Brussels 1090, Belgium

<sup>3</sup>Department of Intensive Care Medicine, UZ Brussel, Laarbeeklaan 101, Brussels 1090, Belgium

<sup>4</sup>Chair of Al and Digital Medicine, University of Mons, Place du Parc 20, Mons 7000, Belgium

# Received: 15 May 2024 / Accepted: 15 April 2025 Published online: 25 April 2025

#### References

- DiMatteo MR, Haskard-Zolnierek KB, Martin LR. Improving patient adherence: a three-factor model to guide practice. Health Psychol Rev. 2012;6(1):74–91.
- Gallagher RA, Porter S, Monuteaux MC, Stack AM. Unscheduled return visits to the emergency department: the impact of Language. Pediatr Emerg Care. 2013;29(5):579–83.
- 3. Krishel S, Baraff LJ. Effect of emergency department information on patient satisfaction. Ann Emerg Med. 1993;22(3):568–72.
- Al-Harthy N, Sudersanadas K, Al-Mutairi M, Vasudevan S, Bin Saleh G, Al-Mutairi M, et al. Efficacy of patient discharge instructions: A pointer toward caregiver friendly communication methods from pediatric emergency personnel. J Fam Community Med. 2016;23(3):155.
- Weiss BD, Schwartzberg JG, Association AM. Health literacy and patient safety: help patients understand: manual for clinicians [Internet]. AMA Foundation; 2007. (Health literacy and patient safety: help patients understand: manual for clinicians). Available from: https://books.google.be/books?id=quJ aYgEACAAJ
- 6. Kitching JB. Patient information Leaflets the state of the Art. J R Soc Med. 1990;83(5):298–300.
- Williams DM, Counselman FL, Caggiano CD. Emergency department discharge instructions and patient literacy: A problem of disparity. Am J Emerg Med. 1996;14(1):19–22.
- Clarke C, Friedman SM, Shi K, Arenovich T, Monzon J, Culligan C. Emergency department discharge instructions comprehension and compliance study. CJEM. 2005;7(01):5–11.
- Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR. The prevalence of limited health literacy. J Gen Intern Med. 2005;20(2):175–84.

- Hoek AE, Anker SCP, Van Beeck EF, Burdorf A, Rood PPM, Haagsma JA. Patient discharge instructions in the emergency department and their effects on comprehension and recall of discharge instructions: A systematic review and Meta-analysis. Ann Emerg Med. 2020;75(3):435–44.
- Chacon D, Kissoon N, Rich S. Education attainment level of caregivers versus readability level of written instructions in a pediatric emergency department. Pediatr Emerg Care. 1994;10(3):144–9.
- Spandorfer J, Karras D, Hughes L, Caputo C. Comprehension of discharge instructions by patients in an urban emergency department. Ann Emerg Med. 1995;25(1):71–4.
- DeSai C, Janowiak K, Secheli B, Phelps E, McDonald S, Reed G, et al. Empowering patients: simplifying discharge instructions. BMJ Open Qual. 2021;10(3):e001419.
- Hutchinson CL, Curtis K, McCloughen A, Qian S, Yu P, Fethney J. Identifying return visits to the emergency department: A multi-centre study. Australasian Emerg Care. 2021;24(1):34–42.
- Ngai KM, Grudzen CR, Lee R, Tong VY, Richardson LD, Fernandez A. The association between limited english proficiency and unplanned emergency department revisit within 72 hours. Ann Emerg Med. 2016;68(2):213–21.
- Alshahrani M, Katbi F, Bahamdan Y, Alsaihati A, Alsubaie A, Althawadi D, et al. Frequency, causes, and outcomes of return visits to the emergency department within 72 hours: A retrospective observational study. JMDH. 2020;13:2003–10.
- Jack BW, Paasche-Orlow MK, Mitchell SM, Forsythe S, Martin J. An Overview of the Re-Engineered Discharge (RED) Toolkit. Rockville, MD: Agency for Healthcare Research and Quality; 2013 Mar. Report No.: 12(13)-0084.
- Karabacak M, Margetis K. Embracing Large Language Models for Medical Applications: Opportunities and Challenges. Cureus [Internet]. 2023 May 21 [cited 2023 Dec 15]; Available from: https://www.cureus.com/articles/14979 7-embracing-large-language-models-for-medical-applications-opportunitie s-and-challenges
- Bui D, Nakamura C, Bray BE, Zeng-Treitler Q. Automated illustration of patients instructions. AMIA Annu Symp Proc. 2012;2012:1158–67.
- 20. Bradshaw JC. The ChatGPT era: artificial intelligence in emergency medicine. Ann Emerg Med. 2023;81(6):764–5.
- Hastings RS, Powers RD. Abdominal pain in the ED: a 35 year retrospective. Am J Emerg Med. 2011;29(7):711–6.
- Marasco G, Verardi FM, Eusebi LH, Guarino S, Romiti A, Vestito A, et al. Diagnostic imaging for acute abdominal pain in an emergency department in Italy. Intern Emerg Med. 2019;14(7):1147–53.
- Carlucci M, Beneduce AA, Fiorentini G, Burtulo G. Nonspecific Abdominal Pain. In: Agresta F, Campanile FC, Anania G, Bergamini C, editors. Emergency Laparoscopy [Internet]. Cham: Springer International Publishing; 2016 [cited 2023 Dec 15]. pp. 73–8. Available from: http://link.springer.com/https://doi.or g/10.1007/978-3-319-29620-3\_6
- Hoseininejad SM, Jahed R, Sazgar M, Jahanian F, Mousavi SJ, Montazer SH, et al. One-Month Follow-Up of patients with unspecified abdominal pain referring to the emergency department; a cohort study. Arch Acad Emerg Med. 2019;7(1):e44.
- Boendermaker AE, Coolsma CW, Emous M, Ter Avest E. Efficacy of scheduled return visits for emergency department patients with non-specific abdominal pain. Emerg Med J. 2018;35(8):499–506.
- 26. Halsey-Nichols M, McCoin N. Abdominal pain in the emergency department. Emerg Med Clin North Am. 2021;39(4):703–17.
- 27. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. Lancet. 2017;389(10070):736–47.
- Balagué F, Mannion AF, Pellisé F, Cedraschi C. Non-specific low back pain. Lancet. 2012;379(9814):482–91.
- Van Wambeke P, Desomer A, Ailliet L, Berquin A, Demoulin C, Depreitere B et al. Low back pain and radicular pain [Internet]. BE: KCE = Federaal Kenniscentrum voor de Gezondheidszorg = Centre Fédéral d'Expertise des Soins de Santé = Belgian Health Care Knowledge Centre; 2017 [cited 2023 Dec 15]. 160 p. Available from: https://doi.org/10.57598/R287C
- Megalla M, Ogedegbe C, Sanders AM, Cox N, DiSanto T, Johnson H et al. Factors Associated With Repeat Emergency Department Visits for Low Back Pain. Cureus [Internet]. 2022 Feb 4 [cited 2023 Dec 15]; Available from: https://ww w.cureus.com/articles/78096-factors-associated-with-repeat-emergency-dep artment-visits-for-low-back-pain
- Ginsberg Z, Ghaith S, Pollock JR, Hwang AS, Buckner-Petty SA, Campbell RL, et al. Relationship between pain management modality and return rates for lower back pain in the emergency department. J Emerg Med. 2021;61(1):49–54.

- Saragiotto BT, Machado GC, Ferreira ML, Pinheiro MB, Abdel Shaheed C, Maher CG. Paracetamol for low back pain. Cochrane Back and Neck Group, editor. Cochrane Database of Systematic Reviews [Internet]. 2016 Jun 6 [cited 2023 Dec 15];2019(1). Available from: https://doi.org/10.1002/14651858.CD01 2230
- Enthoven WT, Roelofs PD, Deyo RA, Van Tulder MW, Koes BW. Non-steroidal anti-inflammatory drugs for chronic low back pain. Cochrane Back and Neck Group, editor. Cochrane Database of Systematic Reviews [Internet]. 2016 Feb 10 [cited 2023 Dec 15];2016(8). Available from: https://doi.org/10.1002/14651 858.CD012087
- Jones CMP, Day RO, Koes BW, Latimer J, Maher CG, McLachlan AJ, et al. Opioid analgesia for acute low back pain and neck pain (the OPAL trial): a randomised placebo-controlled trial. Lancet. 2023;402(10398):304–12.
- Van De Maat JS, Van Klink D, Den Hartogh-Griffioen A, Schmidt-Cnossen E, Rippen H, Hoek A, et al. Development and evaluation of a hospital discharge information package to empower parents in caring for a child with a fever. BMJ Open. 2018;8(8):e021697.
- De Vos-Kerkhof E, Geurts DHF, Steyerberg EW, Lakhanpaul M, Moll HA, Oostenbrink R. Characteristics of revisits of children at risk for serious infections in pediatric emergency care. Eur J Pediatr. 2018;177(4):617–24.
- Curran JA, Murphy A, Burns E, Plint A, Taljaard M, MacPhee S, et al. Essential content for discharge instructions in pediatric emergency care: A Delphi study. Pediatr Emer Care. 2018;34(5):339–43.
- Schmitt BD. Fever phobia: misconceptions of parents about fevers. Am J Dis Child. 1980;134(2):176.
- Crocetti M, Moghbeli N, Serwint J. Fever phobia revisited: have parental misconceptions about fever changed in 20 years?? Pediatrics. 2001;107(6):1241–6.
- 40. Betz MG, Grunfeld AF.???Fever phobia??? In the emergency department: a survey of children??s caregivers. Eur J Emerg Med. 2006;13(3):129–33.
- Pantell RH, Roberts KB, Adams WG, Dreyer BP, Kuppermann N, O'Leary ST, et al. Clinical practice guideline: evaluation and management of Well-Appearing febrile infants 8 to 60 days old. Pediatrics. 2021;148(2):e2021052228.
- Lee P, Bubeck S, Petro J, Benefits. Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. Drazen JM, Kohane IS, Leong TY, editors. N Engl J Med. 2023;388(13):1233–9.
- Deepl. Deepl Translator [Internet]. Deepl; Available from: https://www.deepl.c om/translator
- 44. Rothrock SG, Rothrock AN, Swetland SB, Pagane M, Isaak SA, Romney J, et al. Quality, trustworthiness, readability, and accuracy of medical information regarding common pediatric emergency Medicine-Related complaints on the web. J Emerg Med. 2019;57(4):469–77.
- Daraz L, Morrow AS, Ponce OJ, Farah W, Katabi A, Majzoub A, et al. Readability of online health information: A Meta-Narrative systematic review. Am J Med Qual. 2018;33(5):487–92.
- WebFX [Internet]. [cited 2024 Apr 19]. Readability Test. Available from: https:// www.webfx.com/tools/read-able/
- Seabold S, Perktold J, Statsmodels: Econometric and Statistical Modeling with Python. In: Proceedings of the 9th Python in Science Conference [Internet]., Austin T. 2010 [cited 2023 Apr 14]. pp. 92–6. Available from: https://conferenc e.scipy.org/proceedings/scipy2010/seabold.html
- 48. Waskom M. Seaborn: statistical data visualization. JOSS. 2021;6(60):3021.

- Ziesenitz Z, Erb, Van Den Anker. O-23 Ibuprofen in infants younger than 6 months: what is the efficacy and safety profile? Arch Dis Child. 2017;102(10):A11.1-A11.
- 50. Herzog LW, Coyne LJ. What is fever?? Normal temperature in infants less than 3 months old. Clin Pediatr (Phila). 1993;32(3):142–6.
- Dang R, Schroeder AR, Weng Y, Wang ME, Patel AI. A Cross-sectional study characterizing pediatric temperature percentiles in children at Well-Child visits. Acad Pediatr. 2023;23(2):287–95.
- 52. Jindal P, MacDermid J. Assessing reading levels of health information: uses and limitations of Flesch formula. Educ Health. 2017;30(1):84.
- Vandeghinste V, Bulté B. Linguistic proxies of readability: comparing easy-to-read and regular newspaper Dutch. Comput Linguistics Neth J. 2019;9:81–100.
- 54. van Oosten P, Tanghe D, Hoste V. Towards an Improved Methodology for Automated Readability Prediction.
- Stossel LM, Segar N, Gliatto P, Fallar R, Karani R. Readability of patient education materials available at the point of care. J GEN INTERN MED. 2012;27(9):1165–70.
- Beaunoyer E, Arsenault M, Lomanowska AM, Guitton MJ. Understanding online health information: evaluation, tools, and strategies. Patient Educ Couns. 2017;100(2):183–9.
- Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthc (Basel). 2023;11(6):887.
- Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med. 2023;183(6):589.
- Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst. 2023;47(1):33.
- Bommasani R, Liang P, Lee T. Holistic evaluation of Language models. Ann N Y Acad Sci. 2023;1525(1):140–6.
- Papineni K, Roukos S, Ward T, Zhu WJ. BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02 [Internet]. Philadelphia, Pennsylvania: Association for Computational Linguistics; 2001 [cited 2025 Feb 14]. p. 311. Available from: http://portal.acm.org/citation.cfm?doid=1073083. 1073135
- Zhang T, Kishore V, Wu F, Weinberger KQ, Artzi Y, BERTScore. Evaluating Text Generation with BERT [Internet]. arXiv; 2020 [cited 2025 Feb 14]. Available from: http://arxiv.org/abs/1904.09675
- 63. Regulation EU 2024/1689 EN. EUR-Lex [Internet]. [cited 2025 Mar 18]. Available from: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
- Congress US. Patient Protection and Affordable Care Act [Internet]. 2010. Available from: https://www.govinfo.gov/content/pkg/PLAW-111publ148/pd f/PLAW-111publ148.pdf

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.