

Utilizing ChatGPT for Streamlined Chart Review in Wound Care

Rebecca Friedman BA^{1*}, Rebecca Lisk BA^{1*}, Katherine Cordero-Bermudez MD², Sofia Ghani¹, Soniya Singh², Brian M. Gillette PhD², Scott A. Gorenstein MD², Ernest S. Chiu MD FACS¹

1. Hansjörg Wyss Department of Plastic Surgery, New York University Langone Health, New York, New York, USA 2. Department of Surgery NYU Langone Hospital – Long Island, Mineola, New York, USA



BACKGROUND

Traditional chart review in wound care research is increasingly challenged by a growing volume of patient data and extensive variables impacting wound outcomes. The emergence of **Natural Language Processing (NLP)** software such as **ChatGPT** presents a unique opportunity to automate the data extraction process. The capabilities of **ChatGPT 4.0**, hosted by our medical center’s secure, private Azure OpenAI service were evaluated, to automatically extract and process variables from patient charts following **sacral wound** visits.

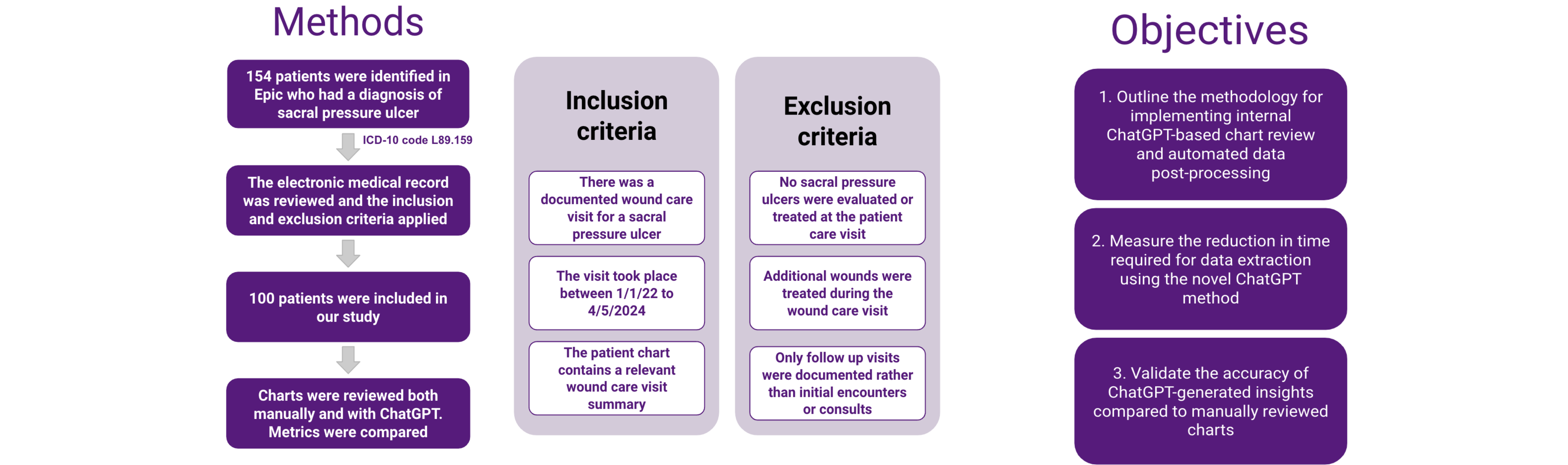
The purpose of this study is to harness ChatGPT’s potential to **revolutionize chart review** through improved **data retrieval accuracy** and **efficiency**.

METHODS

ChatGPT and a **Python script** were integrated into the existing chart review process at our institution. Patients from **NYU Langone** in Manhattan and Long Island were evaluated. Notes from visits for patients with **sacral wounds** were examined. We **manually extracted** multiple variables related to wound care.



Each chart was extracted by **two different study members**, as well as run through the integrated **ChatGPT** platform for automated extraction. Accuracy of extracted variables were compared **between each human extractor**, each **human versus ChatGPT**, and **ChatGPT against a second iteration of ChatGPT extraction**. Metrics include **time taken for review**, **comparative accuracy** of extracted information across and between providers, and **inter-rater reliability** for extraction.



RESULTS

Variable	Accuracy
Age	0.987 (0.977 - 0.995)
BMI (kg/m^2)	0.987 (0.977 - 0.995)
Blood pressure	0.930 (0.890 - 0.970)
Temperature	0.936 (0.918 - 0.954)
Current infections	0.795 (0.772 - 0.818)
Years since quitting smoking	0.936 (0.918 - 0.954)
HbA1c (%)	0.987 (0.977 - 0.995)
WBC (10^9/L)	0.968 (0.957 - 0.979)
Glucose (mmol/L)	0.968 (0.957 - 0.979)
Sodium (mmol/L)	0.974 (0.963 - 0.985)
Potassium (mmol/L)	0.981 (0.970 - 0.991)
CO2 (mmol/L)	0.974 (0.963 - 0.985)
BUN (mg/dL)	0.968 (0.957 - 0.979)
Creatinine (mg/dL)	0.955 (0.941 - 0.970)
C-reactive protein (mg/L)	0.968 (0.957 - 0.979)
Vascular function and ischemia history	0.581 (0.556 - 0.606)
Mobility status	0.810 (0.786 - 0.834)
Spinal cord injury history	0.949 (0.933 - 0.964)
Wound area (cm^2)	0.904 (0.880 - 0.927)
Autoimmune disorders	0.828 (0.804 - 0.852)
Malignancies	0.955 (0.941 - 0.970)
Neurologic disorders	0.731 (0.705 - 0.756)
Current steroid use	0.841 (0.817 - 0.864)

	Human 2	ChatGPT
Human 1	94.55%	95.09%
ChatGPT	96.44%	97.38%

Table 1: Accuracy comparison of chart review results between pairs of reviewers: Human 1 vs. Human 2, Human 1 vs. ChatGPT, Human 2 vs. ChatGPT, and ChatGPT vs. ChatGPT.

Accuracy	Note-writer 1	Note-writer 2	Note-writer 3	Note-writer 4
ChatGPT Review vs. Human	96.12%	93.31%	94.94%	94.77%

Table 3: Average accuracy of ChatGPT chart review vs. manual review by a human for patient encounter notes written by four different physicians

ChatGPT wound descriptions
Patient has a stage 3 sacral pressure ulcer with surrounding erythema and edema, but no odor, discharge, or signs of infection.
Sacral wound 1x3cm with granulating base and fibrinous border.
Sacral wound with 100% healthy granulation tissue and no signs of cellulitis, odor, or excessive drainage.
Stage 4 sacral pressure injury with palpable bone and undermining present, characterized by granulating wound bed, and absence of cellulitis, odor or drainage.
Sacral wound has 100% healthy granulation tissue with no cellulitis, odor or drainage.
The primary wound is a 3x5 cm Stage III granulating wound in the right inner groin.
Large Stage IV sacral pressure ulcer, is present, with wound base down to bone and a size of 20cm x 12cm x 8cm . It exudes serosanguinous fluid and shows macerated wound edges.
The unstageable pressure ulcer is located on the sacrum, has dimensions of 7.5cm by 6.5cm with a dry leathery base down to periosteum.
There is a full thickness ulcer on the sacrum to the left of the natal cleft approximately 2cm x 2cm x 1cm with undermining from 4-7 o'clock approximately 4cm.
The patient's primary wound is a sacral pressure ulcer with hypergranulation tissue.
There is a partial thickness wound on the sacrum which appears granular with no exposed bone, tendon, cellulitis, or purulence.

Table 4: Accuracy achieved by ChatGPT for various chosen variables
Table 5: Wound description summaries compiled by ChatGPT using descriptive text in the chart (not images)

RESULTS

ChatGPT achieved a 96.13% accuracy rate compared to manual chart review conducted by four human reviewers and identified an additional 1.40% of variables that were missed by manual reviewers. Furthermore, the average time per chart review decreased from **7.56 minutes** with the manual method to **1.03 minutes** using ChatGPT. Across four different note writers, similarity between reviews by ChatGPT versus a human was consistent, ranging from 93.31% to 96.12%. ChatGPT was also able to **synthesize accurate summaries** based on patient history, including wound descriptions, prognoses, and risk assessments. An example of a **synthesized summary statement**: The patient has a superficial full thickness ulcer of the sacrum with periwound xerotic scaling and tissue pressure injury, but no exposed bone or tendon, cellulitis, or purulence.

CONCLUSIONS

- The study highlights the potential of ChatGPT to enhance the speed and precision of chart review processes in the context of both clinical care and wound care research
- This offers valuable implications for the integration of artificial intelligence in **healthcare workflows**.

Future Directions

We envision utilization of this powerful tool enhancing the **clinical setting**, allowing physicians to quickly and accurately gather necessary patient data before **examining their patients**, keeping them informed of **relevant history, labs, medications, and a variety of other variables** of the physician's choice in an extremely quick and efficient manner. We also intend to use ChatGPT to greatly enhance our ability to quickly and accurately perform **large scale chart reviews**, allowing researchers to access a **large pool of patient data** with a **fraction of the time and effort**. We expect this to greatly **reduce the burn-out** associated with retrospective research and lead to more robust research being performed in the future.

Contact: Rebecca.friedman@nyulangone.org