FROM IMAGE TO INSTRUCTION: EVALUATING CHATGPT'S ABILITY TO GUIDE MEDICATION USE

Thien Nguyen, Yusef Yassin, Krishna Panchal, Katie Nault, PharmD, Timothy Aungst, PharmD | MCPHS University – W/M, Worcester, MA

BACKGROUND

With the advent of image recognition capabilities within AI programs, patients can now upload a medication image and instantly receive instructions for use. It is unclear the accuracy level of these instructions when compared with those provided by the manufacturer.

Our **OBJECTIVE** is to evaluate the accuracy and safety of medication instructions for use generated by ChatGPT answers from self-taken drug images versus the manufacturer instructions across several different routes of administration.

METHODOLOGY

1 **Twelve medications** were **photographed** three different ways, including by themselves, with packaging, or packaging alone.

|			
Advair Diskus	Ajovy	Aldendronate	Epipen
Flonase	Humira	QVAR Redihaler	Nicorette Gum
Spiriva Handihaler	Symbicort	Voltaren	Zaditor

2 **Images** were uploaded to **ChatGPT4** (Version Jan 2022) in September-October 2023 to generate instructions with the prompt "How do I take this medication."

3 **Responses** were collected, and the text was prepared to be **compared** with the PI from DailyMed. Text similarity was tested based on Cosine Similarity under Count Vectorization and TF-IDF Vectorization providing a **similarity percentage** to either the medication instruction for use (IFU) or MedGuide.

4 The ChatGPT generated medication **instructions** were then **clinically evaluated** for their accuracy based on meeting acceptable patient instructions.
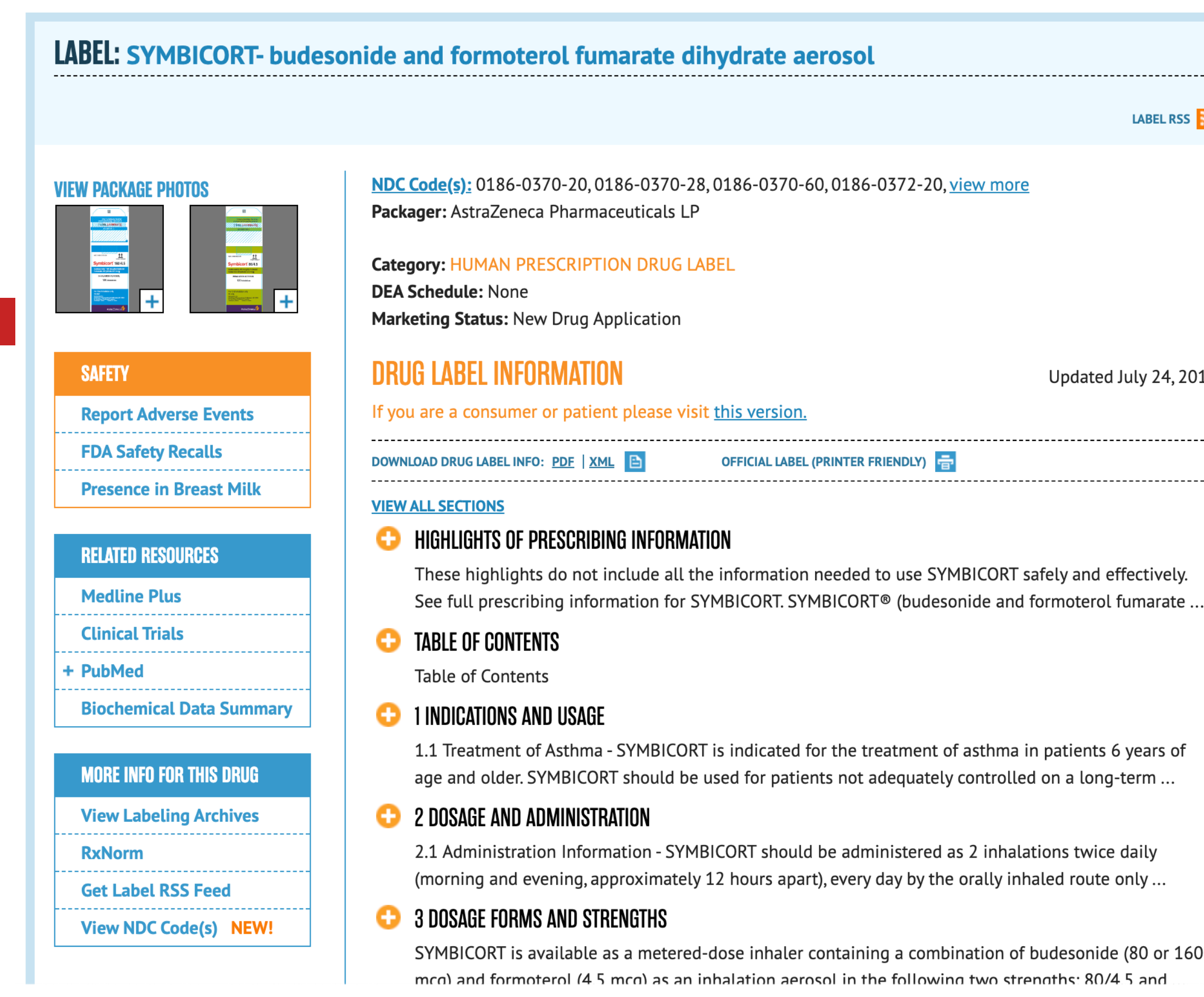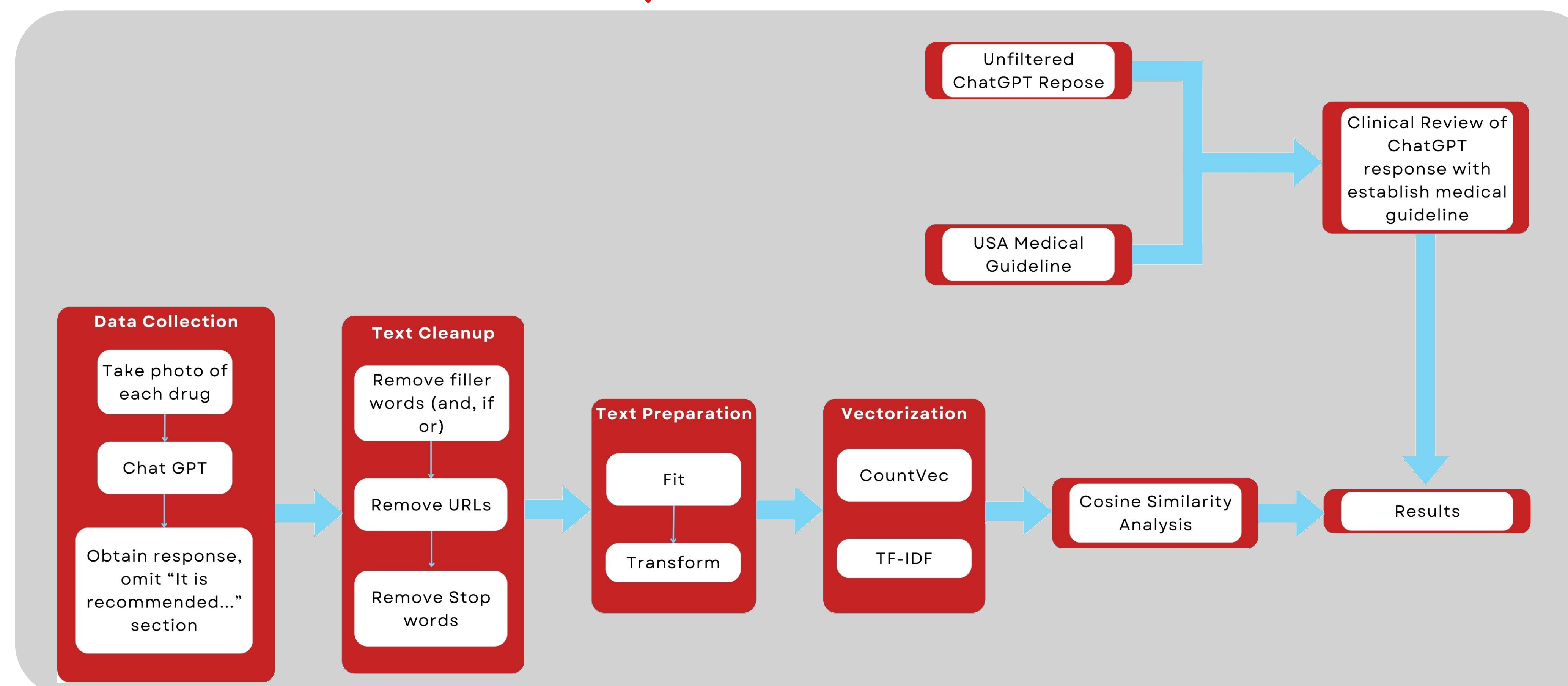
SYMBICORT IMAGE
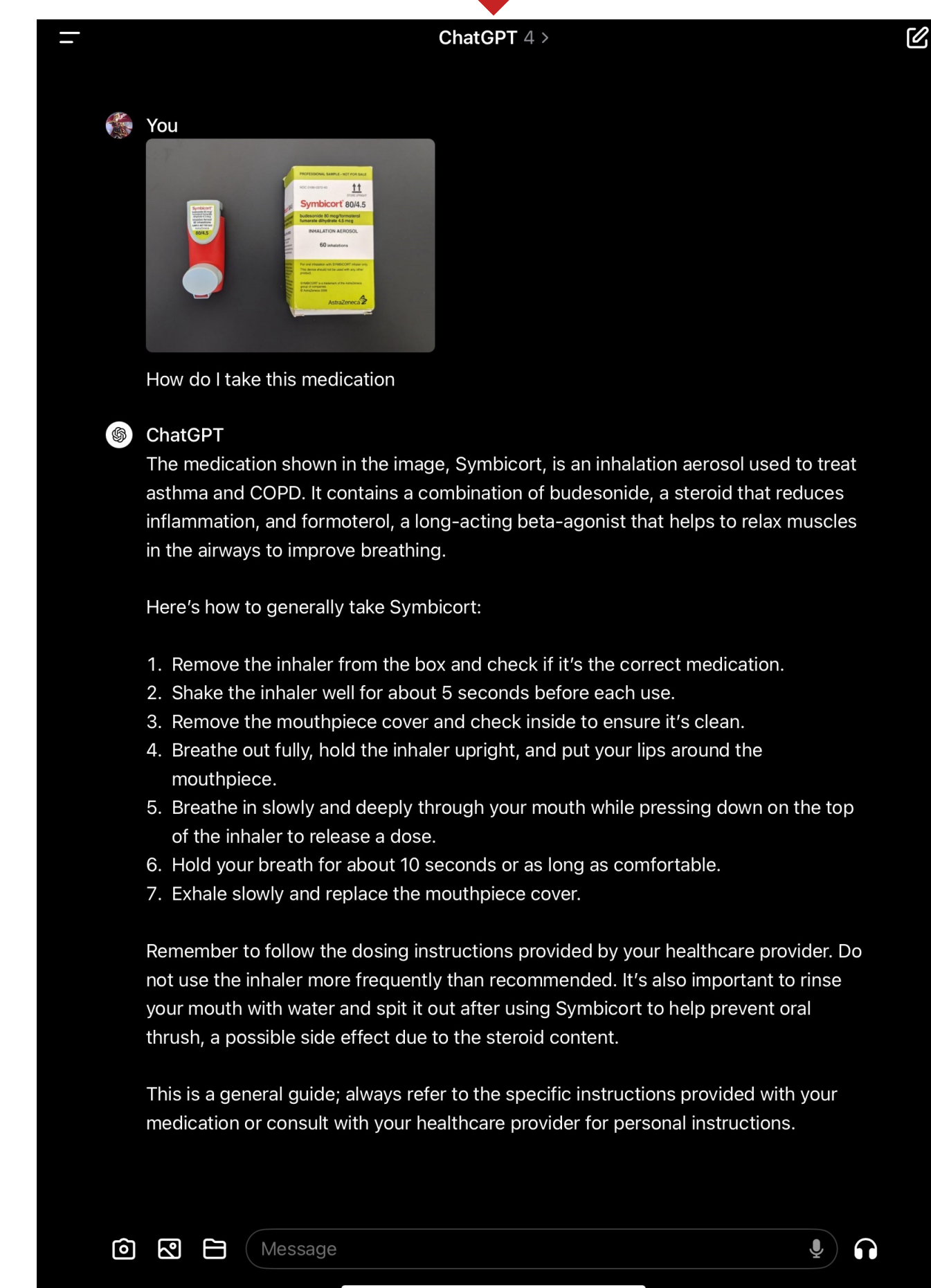
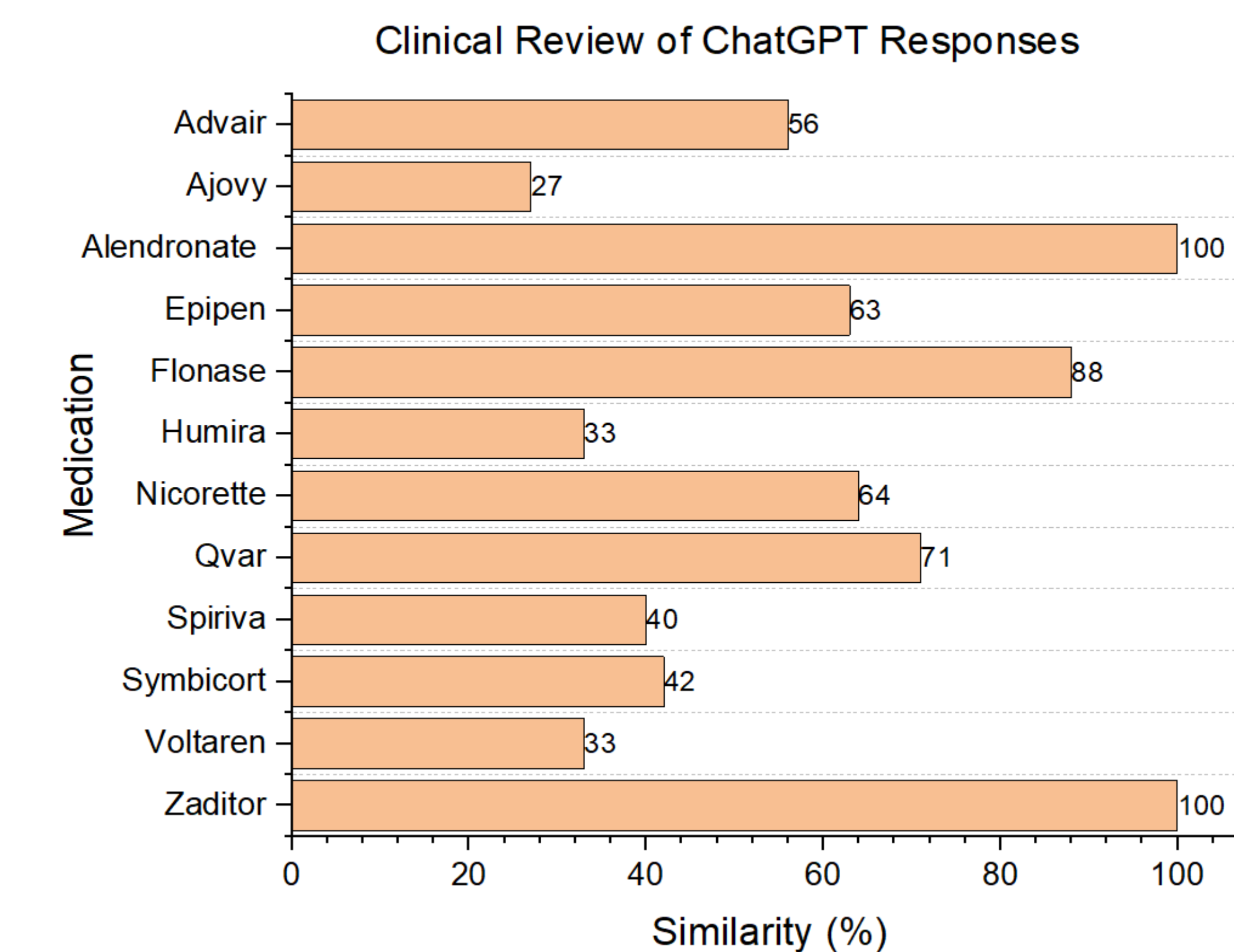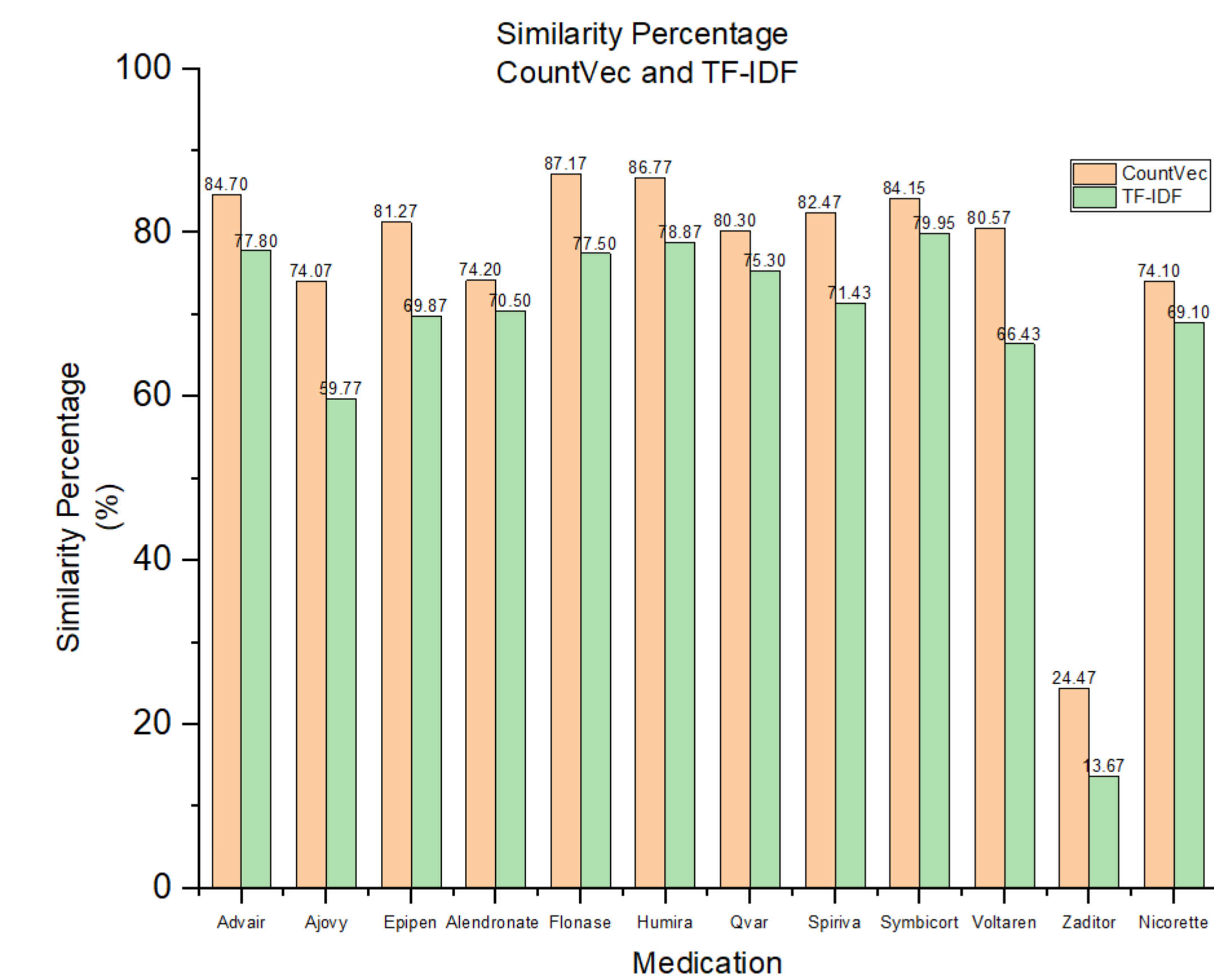**CHATGPT INSTRUCTIONS FOR SYMBICORT**

SYMBICORT PACKAGE INSERT (US NIH NLM)

RESULTS

ChatGPT identified all 12 medications **correctly** and produced patient text instructions on administration.

Similarity Percentage CountVec and TF-IDF

Clinical Review of ChatGPT Responses

SIMILARITY OF GENAI & MEDGUIDE INFORMATION: Between the two similarity comparison methods, **Count Vectorizer** demonstrated a better **similarity performance** than TF-IDF with an average difference of 8.75%. Average text similarity of **ChatGPT** instructions to IFU and MedGuides was 76.24%.

CLINICAL CONTENT EVALUATION: Clinical review scores were lower, with an avg of 59.75%. **ChatGPT** was found to provide **more** patient **friendly instructions** than the PI. Medications with more complicated routes of administration (e.g. injectables) or more steps were found to have **lower** Count Vectorization match rates and clinical review scores. Medications that were administered orally or with fewer steps to use were found to have **higher** similar scores.

CONCLUSION

ChatGPT has the **potential** to generate step-by-step administration instructions and provides more patient-friendly language than the standard IFU or MedGuide. However, its **accuracy** varies with medication administration complexities and discrepancies may arise from its diverse training resources, leading to lower **similarity scores** on complex instructions.