



Comparing Provider and Artificial Intelligence Chatbot Responses to Breast Reconstruction Patient Questions

Daniel Soroudi, B.S.¹, Aileen Gozali, M.M.¹, Nisha Parmeshwar, M.D.², Jacquelyn Knox, M.D.², Ryan Sadjadi, M.P.H.¹, Scott L. Hansen, M.D.², Merisa Piper, M.D.²

¹University of California San Francisco, School of Medicine,

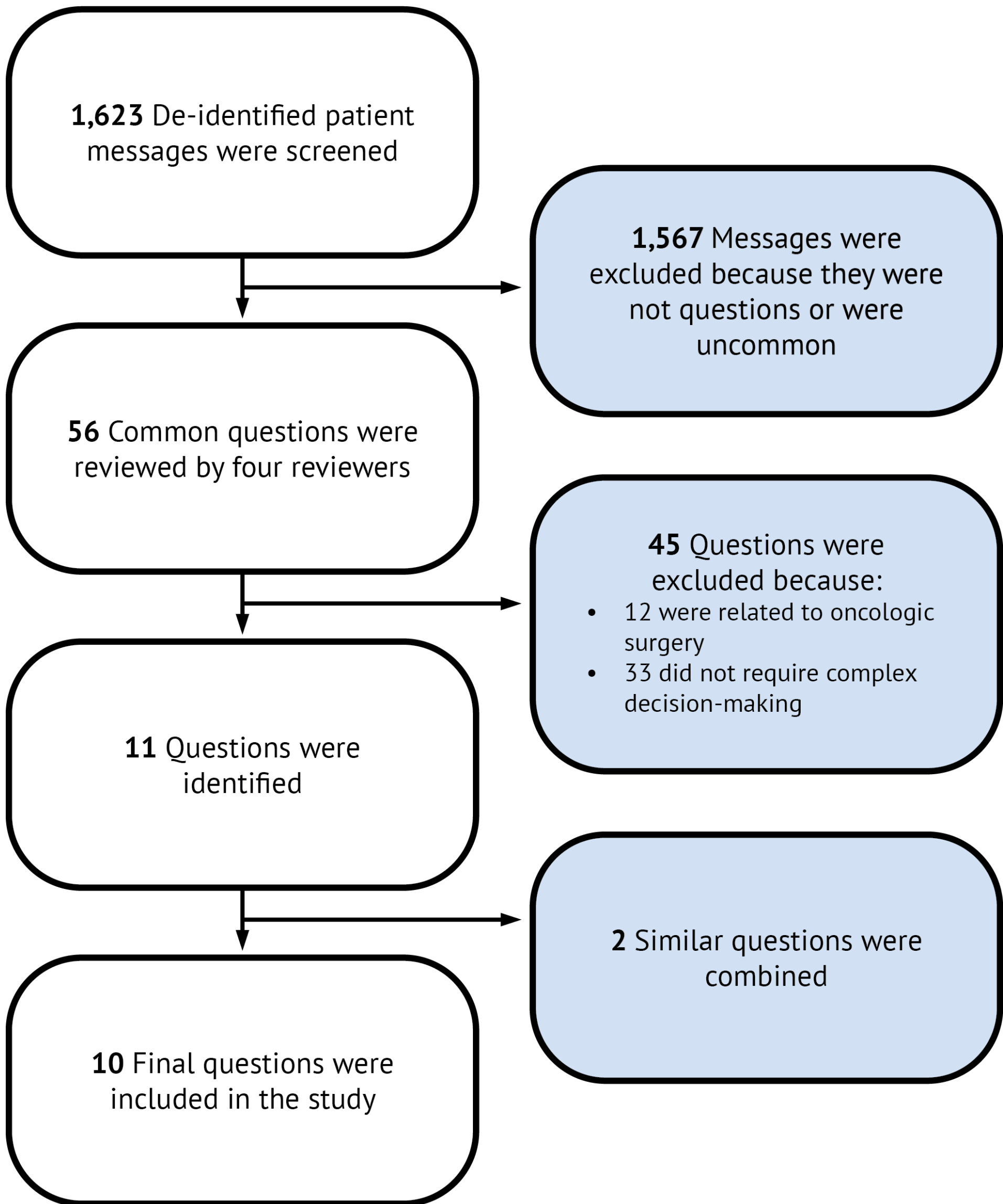
²University of California San Francisco, Department of Surgery, Division of Plastic and Reconstructive Surgery



INTRODUCTION

- Patients express dissatisfaction with information received pre-operatively
- Patient-directed Electronic Health Record (EHR) messaging burdens the physician.
- Tools like ChatGPT can revolutionize patient communication.
- This study assesses ChatGPT's effectiveness in answering breast reconstruction queries.

METHODS



- Questions presented to ChatGPT-3, ChatGPT-4, plastic surgeons, and advanced practice providers (APPs)
- Readability via the Flesch Reading Ease
- Accuracy and Empathy via 1-5 Likert scale

RESULTS

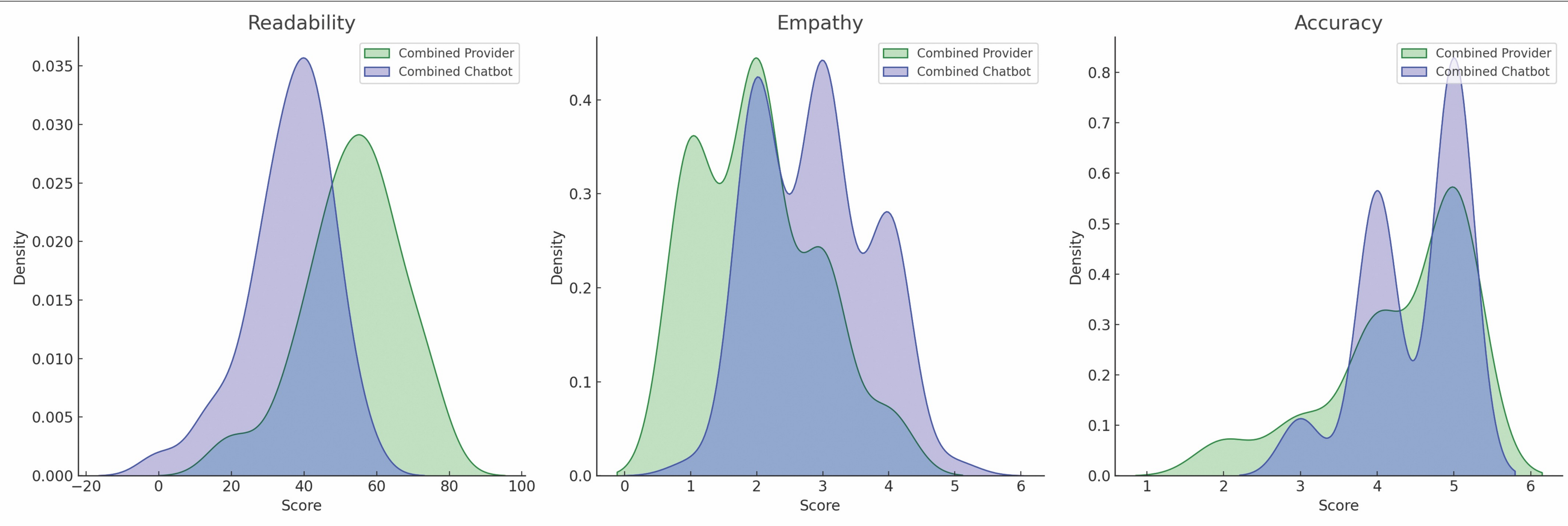
Table 1: Overview of Physician, APP, and Chatbot Response Metrics

Metrics	Physicians	APPs	ChatGPT-3 Full-length	ChatGPT-4 Full-length	ChatGPT-3 Brief	ChatGPT-4 Brief
FRE score	53.2 ± 13.4	53.5 ± 13.5	31.7 ± 10.6	42.8 ± 7.1	31.3 ± 15.5	38.0 ± 9.0
Empathy score	2.1 ± 1.0	2.0 ± 0.8	3.7 ± 0.6	3.4 ± 0.5	2.4 ± 0.6	2.1 ± 0.5
Accuracy score	4.2 ± 0.9	4.4 ± 0.9	4.2 ± 0.6	4.6 ± 0.6	4.4 ± 0.8	4.8 ± 0.4
Word Count	92.0 ± 76.1	74.3 ± 68.2	378.3 ± 88.0	396.9 ± 90.5	63.6 ± 12.6	63.1 ± 21.0
Time to Completion (mins)	23.8 ± 18.4	30.1 ± 18.6	N/A	N/A	N/A	N/A

Table 2: Comparison of ChatGPT-3 and ChatGPT-4 Responses

Metrics	ChatGPT-3	ChatGPT-4	p-value
FRE score	31.5 ± 13.0	40.4 ± 8.3	0.014
Empathy score	3.0 ± 0.9	2.7 ± 0.8	0.108
Accuracy score	4.3 ± 0.7	4.7 ± 0.5	0.004
Word Count	221.0- ± 172.7	230.0 ± 182.8	0.873

Figure 1: Kernel density plot depicting the distribution of Readability, Empathy, and Accuracy scores for Provider and Chatbot Responses.



- Flesch Reading Ease scores for providers were better than AI responses (53.33±13.27 vs. 35.97±11.62, p<.001).
- Empathy scores for providers were lower than AI responses (2.013±0.74 vs. 2.875±0.75, p<.001).
- Accuracy scores were not different between providers and AI responses (4.3 ± 0.9 vs. 4.5 ± 0.6, p=0.170).

CONCLUSIONS

- ChatGPT has potential to enhance patient-provider communication
- Strong in empathy, accuracy, and efficiency of response generation
- Weak in adhering to readability guidelines
- Equitable access to AI technology remains to be addressed
- Future studies should focus on:
 - Optimization of AI tools
 - Exploring the effectiveness of prompting techniques
 - Exploring synergistic effects of AI-provider responses

REFERENCES

1. Islam T, Musthaffa S, Hoong SM, et al. Development and evaluation of a sustainable video health education program for newly diagnosed breast cancer patients in Malaysia. *Support Care Cancer*. 2021;29(5):2631-2638.
2. Sherman KA, Shaw LKE, Winch CJ, et al. Reducing Decisional Conflict and Enhancing Satisfaction with Information among Women Considering Breast Reconstruction following Mastectomy: Results from the BRECONDA Randomized Controlled Trial. *Plast Reconstr Surg*. 2016;138(4):592e-602e.
3. Lee BT, Chen C, Yueh JH, Nguyen MD, Lin SJ, Tobias AM. Computer-based learning module increases shared decision making in breast reconstruction. *Ann Surg Oncol*. 2010;17(3):738-743.
4. Zhong T, Hu J, Bagher S, et al. Decision regret following breast reconstruction: The role of self-efficacy and satisfaction with information in the preoperative period. *Plast Reconstr Surg*. 2013;132(5).
5. Survey: Many U.S. Women Lack Basic Information on Life After Mastectomy | ASPS. Available at: <https://www.plasticsurgery.org/news/press-releases/survey-many-us-women-lack-basic-information-on-life-after-mastectomy>. Accessed July 27, 2023.
6. Readability Scoring System. Available at: <https://readabilityformulas.com/readability-scoring-system.php>. Accessed December 4, 2023.