

T9-P-11

Artificial Intelligence and Hidradenitis Suppurativa:

Validity and reliability of Artificial Intelligence by experts as public sources of information for patients

A.-C. Ezanno¹, C. Fite, C. Jacobzone, J. Gottlieb, C. Pruvost-Balland, G. Gabison, F. Maccari, M. Lamiaux, A. Nassif, P.-A. Becherel, Z. Reguiaïl, J.-L. Perrot, P. Guillem, A.-C. Fougousse and RésoVerneuil
(1)HIA Bégin, Surgery, St Mandé, France

Introduction

- ❖ Hidradenitis Suppurativa (HS)
 - Chronic, inflammatory skin disease.
 - Causes significant pain and discomfort.
 - Patients may hesitate to discuss symptoms.
 - Potential delay in diagnosis and treatment.
- ❖ HS patients increasingly using online resources
- ❖ AI as a Solution
 - AI bridges communication gap.
 - Aims to reduce diagnosis time (currently 8 years)

Materials

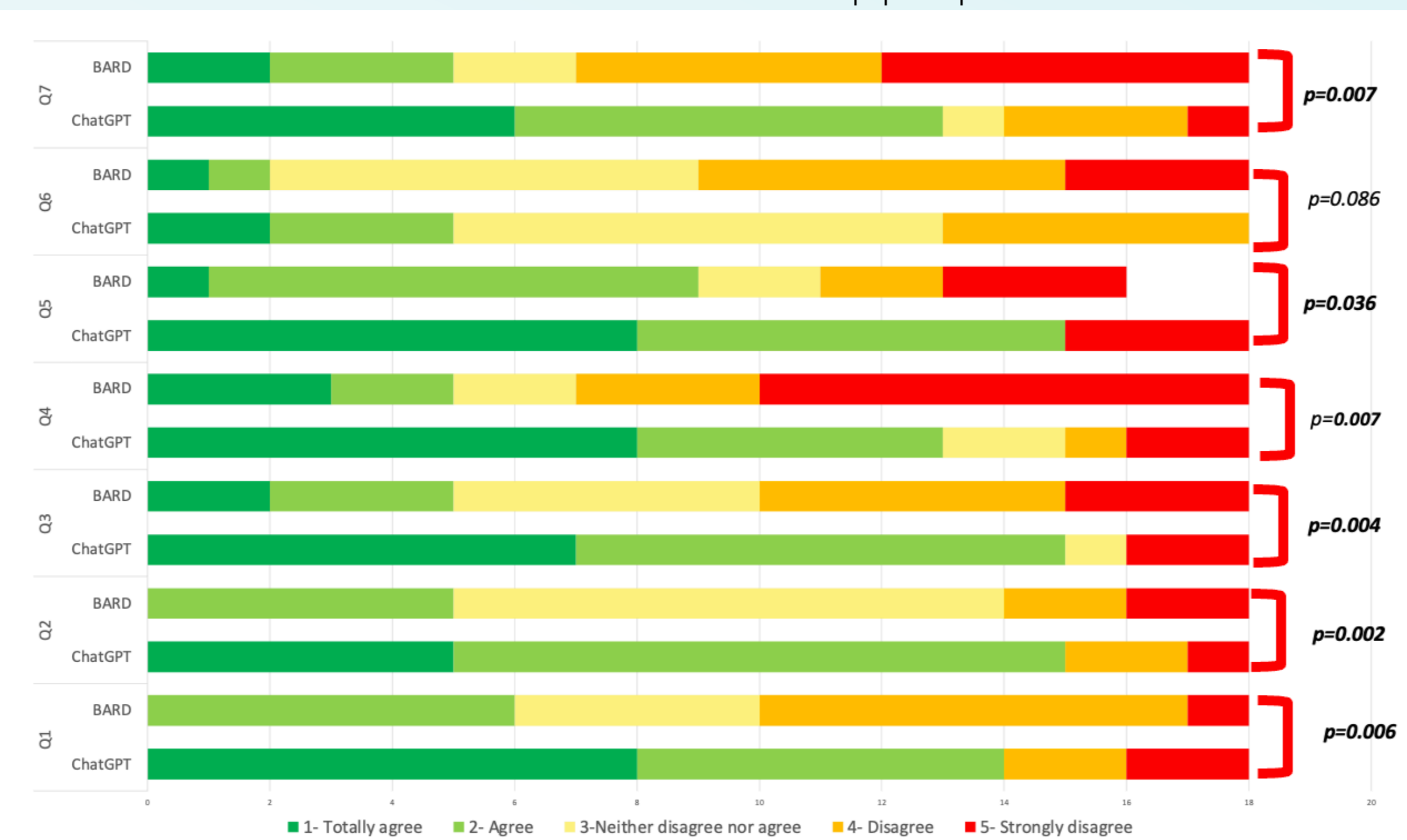
- ❖ Study Objective:
 - Assess and compare validity of 2 AI responses: ChatGPT and Google BARD
 - Questions from potentially/diagnosed HS patients.

Q1: I have abscesses in my armpits and/or perianal folds and/or pubic area all the time. What could it be?
 Q2: I have several abscesses at the same time in the groin crease, armpits, perianal folds... what could it be?
 Q3: What is HS? And is it a serious disease?
 Q4: Is it hereditary?
 Q5: Which doctor do I need to see?
 Q6: What treatments are available?
 Q7: In the meantime, what advice can you give me to improve my situation?

- ❖ Evaluated by a group of HS experts.

Results

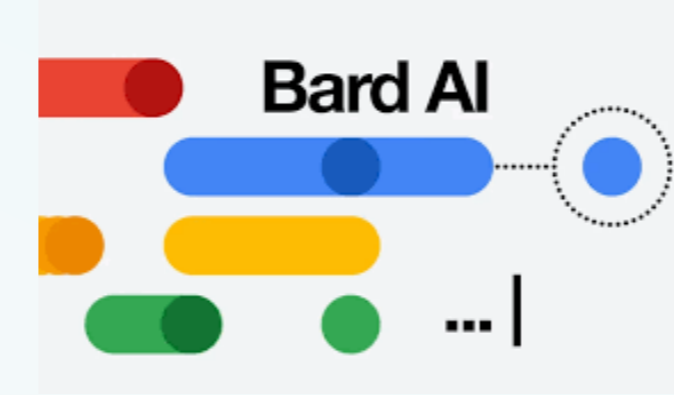
- ❖ ChatGPT provided relatively shorter answers : 228± 48 words versus 254± 77 for BARD
- ❖ Both IAs encouraged users to consult a doctor
- ❖ But neither recommended consult the websites of learned societies
- ❖ ChatGPT answers were deemed as appropriate in 86% and only in 14% for Google BARD



- ❖ Validity of GPT-3.5 responses was significantly higher than that of Google Bard (p < 0.0001)

	Score for ChatGPT 3.5 mean (SD)	Score for BARD, mean (SD)	P value
Q1	2.1±1.4	3.2±1	0.006
Q2	2.1±1.1	3.1±0.9	0.002
Q3	2±1.2	3.2±1.2	0.004
Q4	2.1±1.3	3.6±1.5	0.007
Q5	2.1±1.4	2.8±1.3	0.036
Q6	2.0±0.9	3.5±1	0.086
Q7	2.2±1.2	3.6±1.4	0.007

- ❖ The majority of ChatGPT responses were significantly deemed suitable when comparing to Google BARD
- ❖ Limitation: AI chatbot are not validated medical devices



Results

- ❖ **Validity Comparison** GPT-3.5 vs. Google Bard:
 - GPT-3.5 responses significantly more valid than Bard (p = .008)
- ❖ **Patient Understanding and Efficiency**
 - ChatGPT responses deemed very helpful or quite helpful, especially for patients unaware of their HS diagnosis
 - Less efficient for treatment-related queries
- ❖ **Diagnostic Credibility**
 - Google Bard: provides less credible information, particularly for diagnostic questions related to HS.
- ❖ **Treatment Insights**
 - AI Performance: No clear superiority in treatment information between ChatGPT and Google Bard.
 - Emphasizing advice for enhancing patients' QOL
- ❖ **Quality of Information?**
- ❖ **ChatGPTs' Advantage:**
 - ChatGPT offers superior information on treatment and advice to enhance patients' quality of life

Conclusion

- ❖ AIs, such as ChatGPT and Bard, show promise in aiding access to medical counseling and potentially reducing diagnostic delays.
- ❖ Important to acknowledge the potential for generating biased misinformation, posing risks to patient care.
- ❖ Caution against replacing medical consultation with AI.