MIDWEST ORTHOPAEDICS | RUSH

AAOS OrthoInfo Provides More Accessible Information Regarding Rotator Cuff Surgery Than ChatGPT

Catherine Hand BS, Camden Bohn BA, Yining Lu MD, Marisa Ulrich MD, Erick Marigi MD, Josh Chang BS, Daanish Khazi-Syed BS, Brian Forsythe MD

BACKGROUND

- Over **61% of Americans** seek health information online, often leading to confusion or anxiety.
- ChatGPT is widely used but may be **too complex** or inaccurate for patients with lower literacy.
- OrthoInfo provides **expert-reviewed patient education**.
- Rotator cuff injuries are common, especially in older adults, making readability essential.

PURPOSE

The purpose of this project is to compare the readability and accuracy of AI-generated responses from ChatGPT versus expert-reviewed content from AAOS OrthoInfo for common questions about rotator cuff injuries.

Hypothesis:
➔ ChatGPT may provide accurate information, but OrthoInfo will be more readable and accessible.

METHODS

- 7 FAQs from OrthoInfo's rotator cuff page posed to ChatGPT-4.
- Responses evaluated for readability using 8 standardized tools (e.g. Flesch-Kincaid, Gunning Fog, SMOG).
- Accuracy rated by 3 orthopedic surgeon reviewers using a 4-point scale.
- t-tests and Pearson correlation used to compare readability and accuracy.

RESULTS

	ChatGPT	AAOS OrthoInfo
Gunning Fog Index	17.6	14.9
Flesch-Kincaid Grade Level	13.7	10.1
Coleman-Liau	15.1	11.5
SMOG Readability Formula	15.4	13.7
FORCAST Readability Formula	12.9	10.8
Fry Graph	14.0	11.3
Raygor Readability Estimate	14.0	11.1
Average Grade level:	14.7	11.9
P-value	0.008*	

Table 1: Average readability measurements among different readability calculators calculated for each of the 7 FAQs derived from OrthoInfo.

*: signifies statistical significance

- ChatGPT responses required college-level education (avg. grade level 14.7) vs. OrthoInfo at 11.9 (**p<0.01**)
- ChatGPT had significantly lower Flesch Reading Ease (25.9) vs OrthoInfo (52.5) (**p<0.01**).
- All 7 readability tools showed ChatGPT was significantly harder to read (see tables 1 & 2)
- Accuracy was high for both; however, ChatGPT scored lower for the question on further damage (mean 3 vs 4, **p<0.05**).

	ChatGPT							
	Subsection/Questions	Flesch Reading Ease Index	Gunning Fog Index	Flesch-Kincaid Grade Level	Coleman-Liau	SMOG Readability Formula	FORCAST Readability Formula	Fry Graph	Raygor Readability Estimate
RC	What is the RC	33.9	18.8	11.5	14.49	16.53	12.46	14	14
	Surgical Repair	24.8	18.03	12.9	16.32	15.49	13.24	14	14
	Causes	26.1	17.93	12.5	16.37	15.3	13.07	14	14
	Further Damage	27.1	16.09	12.1	16.57	14.47	13.03	14	14
	Doctor	33.1	16.34	15.46	11.8	14.63	12.74	14	14
	Healing	17.9	17.78	18.21	13.5	15.18	13.23	14	14
	Surgery	18.5	18.25	13.3	16.9	15.87	12.84	14	14
	Total:	25.9	17.6	13.7	15.1	15.4	12.9	14.0	14.0
		OrthoInfo							
	Subsection/Questions	Flesch Reading Ease Index	Gunning Fog Index	Flesch-Kincaid Grade Level	Coleman-Liau	SMOG Readability Formula	FORCAST Readability Formula	Fry Graph	Raygor Readability Estimate
	What is the RC	49.3	14.68	11.8	10.21	13.43	10.17	11	9
	Surgical Repair	53.1	16.01	8.3	13.13	14.55	11.65	11	12
	Causes	64.3	13.05	8.1	9.2	12.16	9.51	9	9
	Further Damage	56.7	12.96	9	9.4	12.34	10.07	9	9
	Doctor	43.1	16.41	12.1	15.46	14.84	10.97	13	13
	Healing	50.9	18.3	9.1	13.73	15.9	11.43	12	12
	Surgery	49.8	13.1	12.49	9.5	12.46	11.59	14	14
	Total:	52.5	14.9	10.1	11.5	13.7	10.8	11.3	11.1
	Subsection/Questions	P-value							
	What is the RC	<0.05*							
	Surgical Repair	<0.05*							
	Causes	<0.05*							
	Further Damage	<0.05*							
	Doctor	0.6							
	Healing	0.19							
	Surgery	0.05							

Table 2: Readability scores for both ChatGPT and OrthoInfo responses to the FAQs provided by a number of different readability calculators.

*: signifies statistical significance

DISCUSSION

- ChatGPT is accurate but often **requires a higher literacy level** to understand.
- OrthoInfo is **more readable**, making it better suited for diverse patient populations.
- AI tools need refinement to match readability needs and avoid overwhelming users.
- Future development should focus on personalized, adaptive AI for patient education.