PEDIATRIC DENTISTS' PERSPECTIVES ON ACCURACY & HELPFULNESS OF ChatGPT's ORAL HEALTH INFORMATION

Victoria C. Chang, DMD[1], Nawal Dairi, BS[2], David M. Avenetti, DDS, MSD, MPH[1]

[1] Department of Pediatric Dentistry, UIC College of Dentistry, Chicago, IL; [2]Doctor of Medicine in Dentistry Candidate, University of Pennsylvania, Philadelphia, PA

Background

- Within dentistry, the use of artificial intelligence (AI) is becoming increasingly common, with potential to advance patient education/care, diagnostics, clinical decision making, and treatment planning.
- Chatbots that use a large language model (LLM), such as ChatGPT by OpenAI, have the potential to provide meaningful oral health information to patients and caregivers/parents.
- ChatGPT responses generally demonstrate a high level of accuracy, comprehensiveness, reliability and consistency, in alignment with the current literature.
- However, ChatGPT as a patient education tool has reported limitations: responses may be challenging to comprehend, inaccurate or incomplete, and lack references.
- Although ChatGPT is expected to become more useful and effective over time, responses must receive verification and feedback from related healthcare practitioners. False or misleading information can affect patient decision making and their health outcomes, especially those with low oral health literacy.

Objectives

1. Describe Pediatric Dentistry Program Directors' (PDs) usage and familiarity with ChatGPT.
2. Evaluate readability of ChatGPT responses to parents' clinical questions (FAQs) using word count and Flesch-Kincaid (FK) readability tests.
3. Assess whether PDs find ChatGPT responses to be accurate and helpful.

Figure 1. Flesch-Kincaid Reading Ease Score and Reading Grade Level.

Estimated School Grade Completed	Reading Level	Average Number of Words/Sentence	Average Number of Syllables/Word	Flesch-Kincaid Reading Ease Score	Approximate % of Adults Who Can Read at This Level
4th	Very Easy	8 or Fewer	1.23 or Fewer	90-100	93
5th	Easy	11	1.31	80-90	91
6th	Fairly Easy	14	1.39	70-80	88
7th or 8th	Standard	17	1.47	60-70	83
Some high school	Fairly Hard	21	1.55	50-60	54
High school or some college	Hard	25	1.67	30-50	33
College	Very Hard	29 or More	1.92 or More	0-30	4.5

Methods

Study Design: In this cross-sectional survey, clinical questions were selected across four topic domains for inclusion: 1) Access to Care and Prevention, 2) Treatment and Emergencies, 3) Growth and Development, and 4) Dental Materials and Safety. Primary outcomes are accuracy, helpfulness, and readability of ChatGPT responses based on word count and FK readability tests.

Data Collection:
- ChatGPT responses to the 15 FAQs were collected in October 2023 using ChatGPT 3.5, which is a free and publicly accessible version of ChatGPT. Responses to the FAQs were collected once more in April 2024, again using ChatGPT 3.5, as LLMs evolve with time and training.
- Readability Assessment: Microsoft Word's readability statistics of all FAQ responses were collected for: word count, FK reading ease score, and FK grade level.
- The survey was sent to all US Residency Program Directors from the current American Academy of Pediatric Dentistry list-serv. Program directors completed questions about demographics, AI usage and familiarity, and assessed the accuracy and helpfulness of each ChatGPT response.

Results

- Between Feb-May 2024, 28 surveys were completed (36% response rate), and 4 were excluded.
- Respondents were an average of 49 years of age (SD=8.9), 58% white, and 63% female.
- 75% (N=18) of PDs were "*somewhat*" or "*very familiar*" with ChatGPT. Usage of ChatGPT was as follows: 7 "*never*," 13 "*rarely or sometimes*," 4 "*often*," and 0 "*very often*."
- 58% (N=14) of PDs felt patients would use ChatGPT "*often*" or "*very often*" as a source of dental information in the future, but half disagreed or were neutral that ChatGPT would have a positive impact on patient treatment choices.

Table 1. Flesch-Kincaid Readability of ChatGPT Responses to Clinical Questions.

Question	Word Count October 2023	Word Count April 2024	FK Reading Ease October 2023	FK Reading Ease April 2024	FK Grade Level October 2023	FK Grade Level April 2024
Access to Care and Prevention						
When should my child first see a dentist?	363	77	43.6	44.5	13	13.4
How often / how long should my child brush their teeth?	381	80	60.3	61.7	8.6	9.6
What causes cavities?	408	178	38.1	43.5	13.1	11.9
How can I prevent cavities?	452	347	50.9	48.2	10.5	10.9
Why should my child floss?	353	190	49.8	46.9	11.3	11.6
Treatment and Emergencies						
Is it worth fixing baby teeth	282	88	48.7	43.1	11.5	12.7
How can I make my child's toothache go away?	397	295	48.2	38.1	11.5	13.1
How can the dentist treat my anxious child	466	332	43.7	28.5	11.6	14.6
Growth and Development						
When will my child get their baby teeth?	232	79	48.7	47.5	10.8	11.5
When will my child get their adult teeth?	266	74	42.5	45.7	11.4	13
When should my child get braces?	366	128	38.2	33.8	13.1	13.8
Why are my child's teeth coming in crooked?	353	97	46.3	41	12.1	10.9
Dental Materials and Safety						
Is fluoride safe?	192	148	29.2	31.4	14.2	13.4
Are fillings safe?	384	240	38.6	35.3	12.7	13.3
Is nitrous oxide (laughing gas) safe?	350	153	31.1	29.8	13.3	13.8
Is dental sedation safe?	406	211	21.6	24.5	15.1	14.1
Is general anesthesia safe?	413	252	5	12.4	17.3	16.1
Average (SD)	356.7 (74.5)	174.6 (92.0)	40.3 (13.0)	38.6 (11.3)	12.4 (2.0)	12.8 (1.6)
Median (Min – Max)	366 (192 – 466)	153 (74-347)	43.6 (5-60.3)	41 (12.4-61.7)	12.1 (8.6-17.3)	13.1 (9.6-16.1)

- At both timepoints assessed, ChatGPT responses were lengthy, but there was a slight decrease in average word count (356 and 174 words respectively) over the 6-month period *(Table 1)*.

- The average FK reading grade level remained high (12.4 and 12.8), while the FK reading ease remained at a "High School or Some college" level. Only one item was <10th grade reading level *(Table 1)*.

Results

Table 2. Program Directors' Assessment of Item Accuracy.

Question	Both Correct and Comprehensive	Correct but Incomplete/ Inadequate	Some Correct and Incorrect Information	Completely Incorrect
Access to Care and Prevention (N=20)				
When should my child first see a dentist?	16	3	1	0
How often/how long should my child brush their teeth?	14	4	2	0
What causes cavities?	15	3	2	0
How can I prevent cavities?	12	6	2	0
Treatment and Emergencies (N=20)				
Is it worth fixing baby teeth	12	7	1	0
How can I make my child's toothache go away?	5	6	9	0
How can the dentist treat my anxious child	13	5	2	0
Growth and Development (N=20)				
When will my child get their baby teeth?	14	4	2	0
When will my child get their adult teeth?	11	8	1	0
When should my child get braces?	8	7	5	0
Dental Materials and Safety (N=19)				
Is fluoride safe?	8	11	0	0
Are fillings safe?	7	7	5	0
Is nitrous oxide (laughing gas) safe?	9	5	5	0
Is dental sedation safe?	8	8	3	0
Is general anesthesia safe?	10	6	3	0
Average	10.4	6	2.9	0

Table 2 Discussion:
- Generally, PDs felt that ChatGPT responses were both correct and comprehensive. No one felt that any ChatGPT response was completely incorrect.
- "Dental Materials and Safety" as a topic yielded the most variable distribution in perceived accuracy.

Key Finding:
- For common FAQs, responses are generally accurate sources of information, but some may be incomplete or contain some incorrect information.

Table 3. Program Directors' Assessment of Item Helpfulness.

Question	Very Helpful	Somewhat Helpful	Not Very Helpful	Not Helpful at All
Access to Care and Prevention (N=20)				
When should my child first see a dentist?	14	6	0	0
How often & for how long should my child brush their teeth?	15	5	0	0
What causes cavities?	11	9	0	0
How can I prevent cavities?	9	11	0	0
Treatment and Emergencies (N=20)				
Is it worth fixing baby teeth	11	8	1	0
How can I make my child's toothache go away?	7	7	6	0
How can the dentist treat my anxious child	13	7	0	0
Growth and Development (N=20)				
When will my child get their baby teeth?	12	8	0	0
When will my child get their adult teeth?	8	12	0	0
When should my child get braces?	9	8	3	0
Dental Materials and Safety (N=19)				
Is fluoride safe?	8	11	0	0
Are fillings safe?	4	12	3	0
Is nitrous oxide (laughing gas) safe?	6	12	1	0
Is dental sedation safe?	6	12	1	0
Is general anesthesia safe?	7	11	1	0
Average	9	9.5	1.1	0

Table 3 Discussion:
- Generally, PDs felt that the ChatGPT responses were "*very*" or "*somewhat*" helpful.
- No one felt that ChatGPT responses were "*not helpful at all*," and very few responses were deemed to be "*not very helpful*."

Key Finding:
- For common FAQs, responses are generally "*very helpful*" or "*somewhat helpful*" sources of information.

Conclusions and References

1. In general, Pediatric Dentistry PDs found ChatGPT responses to 15 FAQs across four topic domains to be accurate, comprehensive, and helpful.
2. The lengthy responses are a consistent limitation of ChatGPT. They may exceed the comprehension of the average person and therefore may not be helpful to high-risk populations with low oral health literacy.
3. High-risk populations may be more likely to need dental treatment under sedation or general anesthesia, and seek information from the internet, but these ChatGPT responses were at an especially high reading level (as high as 17.3 FK grade level).
4. Though ChatGPT responses are deemed accurate and helpful from the pediatric dentists' perspective, assessing caregivers' perception of the accuracy and helpfulness of the same FAQs would better inform the utility of ChatGPT and shed light on using LLMs as an information source.