Physician Management of Dental Trauma Utilizing Large Language Models

Lauren Eckert, DMD; Daniel Stoeckel, DDS MS, Tony McClure, DDS
Saint Louis University Center for Advanced Dental Education

Introduction

This study was conducted to assess how medical providers in pediatric emergency departments approach traumatic dental injuries (TDIs) when using large language models (LLMs) compared to traditional text guidelines. The aims were to obtain a baseline knowledge of TDI management, and to assess differences in treatment planning and time taken to answer questions between study groups.

Hypothesis: The group using an LLM (ChatGPT) will have faster response times to the quiz questions, but answers will be less accurate than those of the group using the AAPD guidelines. The control group will be the fastest, but least accurate.

Review of Literature

Often referred to as “AI”, LLMs are computer models trained on vast sets of data to output responses based on what words are most likely to come next. While they can quickly summarize and relay information, LLMs have the potential to provide false statements confidently[1]. Current uses in healthcare include automation of repetitive tasks like charting and patient use of AI chatbots to identify symptoms[2].

A 2025 study asked the models to answer T/F questions on dental avulsions and found ChatGPT had >95% accuracy[2]. However, others were less accurate and were not consistent with answers. Another asked the models to review photographs of TDIs and create a treatment plan. ChatGPT again had the most accurate responses[3].

To date, there have been no studies assessing how physicians can use LLMs to determine treatment needs for TDIs. Most research in this area has involved asking the questions directly to the LLM.

Methodology

We conducted a three-arm, randomized, survey-based study with medical providers in a Children’s Emergency Department (ED). Participants included medical students, residents, and fellows, attending physicians, and advanced practice providers.

Each participant answered 12 case-based questions covering tooth avulsion, tooth fracture, and antibiotic indications using either no resources (**No Resources**), a PDF of the IADT reference guide (**PDF**), or ChatGPT (**ChatGPT**). We recorded how many questions each participant answered correctly, incorrectly, or marked "not sure," as well as how long they took to complete the survey and how confident they felt in their TDI knowledge before starting.

Fig 1. Example of case vignette and multiple-choice question

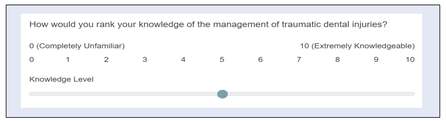

Fig. 2. Self-reported confidence scale

18 participants completed surveys. Due to the small sample size, non-parametric statistical tests (Kruskal-Wallis, Fisher’s) were used to compare ordinal outcomes and analyze correctness rates across groups. Spearman rank-order was used for self-assessment scores and total correct responses.

Results

For all metrics tested, the only outcome to reach significance was a positive correlation between the self-assessment score and % questions correct for No Resource (p=0.003). The pre-survey self-assessment medians (0-10) were: No Resources (6.0), PDF (3.0), ChatGPT (3.3). No Resource had relatively more residents and attendings and ChatGPT had more medical students and APPs.

The areas with the highest % correct answers were primary tooth avulsion and permanent tooth re-implantation timing, storage medium, and splinting. The lowest were antibiotic indications for avulsions and root fractures, the use of Doxycycline, and indications for soft-tissue imaging for tooth fractures.

While not significant, there was increased average accuracy and decreased average time taken when using the ChatGPT vs the PDF. No Resource had the lowest % accurate and fastest completion time.

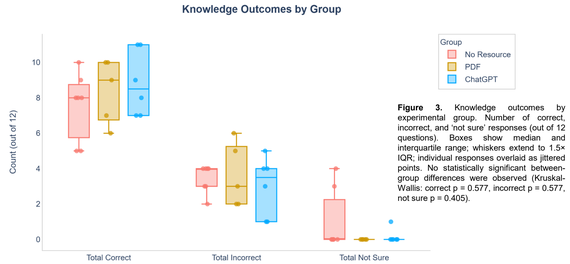

Figure 3. Knowledge outcomes by experimental group. Number of correct, incorrect, and ‘not sure’ responses (out of 12 questions). Boxes show median and interquartile range; whiskers extend to 1.5× IQR; individual responses overlaid as jittered points. No statistically significant between-group differences were observed (Kruskal-Wallis: correct p = 0.577, incorrect p = 0.577, not sure p = 0.405).

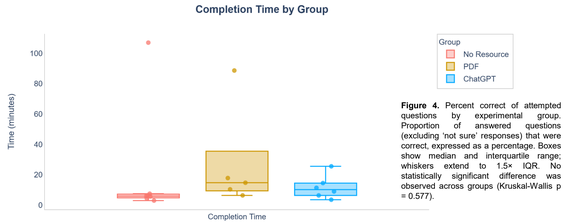

Figure 4. Percent correct of attempted questions by experimental group. Proportion of answered questions (excluding ‘not sure’ responses) that were correct, expressed as a percentage. Boxes show median and interquartile range; whiskers extend to 1.5× IQR. No statistically significant difference was observed across groups (Kruskal-Wallis p = 0.577).

Discussion

This study assessed medical provider baseline knowledge and confidence in TDI treatment planning. It measured clinician speed and accuracy when using different reference materials. Regardless of reference material used, we found gaps in knowledge in multiple areas, particularly antibiotic use. This may indicate a need for a TDI guide geared toward medical providers written by pediatric dentists.

The major limitation of this study is the small sample size. Providers are busy and may not have wanted to use work or leisure time to complete a survey. This may have affected the “completion time” if a provider stepped away during the survey. Additionally, the groups were not evenly distributed by education status. Differing years of experience may have affected how much prior knowledge a participant had.

Future research building on this protocol could include extended response answers to scenarios, collecting confidence scores per question, and assessing different LLMs to compare accuracy.

Conclusion

LLMs may be a useful adjunct for medical providers in the treatment of TDIs, particularly in time-sensitive situations where a consulting dentist is not available. Additional studies are needed to assess how healthcare providers can utilize LLMs to retrieve accurate information quickly and improve patient treatment outcomes.

References

1. Mustuloğlu, Ş., & Deniz, B. P. (2025). Evaluation of Chatbots in the Emergency Management of Avulsion Injuries. *Dental traumatology : official publication of International Association for Dental Traumatology*, *41*(4), 437–444. https://doi.org/10.1111/edt.13041
2. Bajwa, J., Munir, U., Nori, A., & Williams, B. (2021). Artificial intelligence in healthcare: transforming the practice of medicine. Future healthcare journal, 8(2), e188–e194. https://doi.org/10.7861/fhj.2021-0095
3. Çege, E. E., Cömert, H., Akal, N., & Ölmez, A. (2025). Evaluation of the Performance of Artificial Intelligence Based Chatbots in Providing First Aid Information on Dental Trauma According to the ToothSOS Application. Dental traumatology : official publication of International Association for Dental Traumatology. https://doi.org/10.1111/edt.13078