



Integrating ChatGPT into Health Professions Education: A Workshop-Based Approach to Enhance Understanding and Utilization

G. Gershon, D. Kulp, N. Skopicki, M. van Assen, J. Gichoya

DOI: doi.org/10.59450/LFIX6692

PUBLICATION DATE: 08/18/2025

TAKE HOME POINT – A structured, single-session workshop can significantly improve health profession students' comfort and likelihood of using ChatGPT, a tool that is likely to be used increasingly in pre-clinical medical education. This approach could be applied across various health disciplines to enhance learning and foster critical thinking, while promoting interdisciplinary collaboration.

INTRODUCTION

As the demands on the healthcare industry grow more complex and diverse, so too does the need for innovative approaches to training future physicians, nurses, and allied healthcare professionals who are faced with never-ending coursework material and research articles. With the advent of publicly available large language models (LLMs), the integration of artificial intelligence (AI) into medical education has emerged as a powerful tool, with the potential to augment – and ultimately transform – the way students and practitioners learn and hone their skills (Alam et al., 2023; Jeyaraman et al., 2023; Tsang, 2023).

Among the myriad of AI-driven technologies, ChatGPT – Generative Pre-trained Transformer 3.5, an advanced language model developed by OpenAI (San Francisco, CA) – has garnered significant attention for its remarkable capacity to comprehend and generate human-like text. Released in November 2022, ChatGPT's engagement skyrocketed, reaching over 100 million global users in just two months (Roose, 2023; Shivaprakash, 2023; Subbaraman, 2023). ChatGPT-4 is touted as having the capacity to understand and generate text with higher relevance and context sensitivity, as well as the ability to accept multiple input types, including video, voice and non-textual data.

This LLM model has shown promise across numerous applications, including natural language understanding, content generation, and problem-solving (Shorey et al., 2024). In the context of health professions education, ChatGPT represents a groundbreaking advancement, offering educators and learners the methods and tools to revolutionize knowledge acquisition, clinical decision-making, and continuous professional development. Among other applications, the technology has already passed National Board of Medical Examiners (NBME) and United States Medical Licensing Examination (USMLE) Step exams, formulated disease scripts and contemplated differential diagnoses and assisted in scientific manuscript writing (Kung et al., 2023; Mohammad et al., 2023).

The introduction of ChatGPT has also garnered criticism and stirred controversy in academia and in the broader medical community (Kanjee et al., 2023; Palmer, 2023). The current model was only trained on information until September 2021, leaving significant gaps in current evolutions in scientific discovery and thought, particularly in the post-COVID era. While the model correctly identifies commonly cited medical diagnoses and red flag symptoms, it struggles with more advanced analyses of human disease (Duong & Solomon, 2023; Eriksen et al.,

2023; Kanjee et al., 2023). In addition, the diagnostic accuracy of ChatGPT compared to experienced physicians remains an ongoing area of research. Recent studies highlight that while ChatGPT can generate differential diagnoses well, its performance varies based on case complexity and may lack the nuanced clinical reasoning of seasoned physicians (Hirosawa et al., 2023). Mehnen et al. tested ChatGPT's diagnostic performance across 50 clinical cases, including ten rare presentations (Mehnen et al., 2023). The model accurately identified common conditions within its top two differentials, whereas rare diseases required at least eight suggestions to achieve 90% diagnostic coverage. Tan et. al. addressed challenges and risks of using ChatGPT in medicine, most prominently AI "hallucinations," where the system generates seemingly credible but factually incorrect information, including fabricated references (Tan et al., 2024).

In clinical settings, these challenges are compounded by privacy concerns, as optimal performance would require access to sensitive patient data, raising questions about information security and confidentiality. The current absence of clear regulatory frameworks further complicates potential clinical implementation (Mehnen et al., 2023; Tan et al., 2024). Addressing these challenges will be essential to responsibly harness

AI's potential while mitigating risks to patient care, research integrity, and medical education.

Whether ChatGPT can be harnessed to enhance the learning experience of interdisciplinary health education students in collaboration with a traditional educational environment is unknown. Therefore, this study tested the hypothesis that a prospectively designed and innovative educational session on ChatGPT would bolster the immediate and sustained ability of health professions students to understand and utilize ChatGPT and to determine their perception of ChatGPT's utility in their future medical careers.

METHODS

Participants

Students at the Emory School of Medicine, including those from genetic counseling, first and second-year medical students, and first and second-year physician assistant students, provided the cohort for the study. Fifty-nine self-selected students with varying exposure to ChatGPT participated in the workshop, which had the learning objectives listed below.

By the end of this workshop, learners will be able to:

1. Understand the framework of an

LLM like ChatGPT and be able to effectively explain it to a peer.

2. Develop comfort in using ChatGPT within medical contexts, as demonstrated by generating relevant prompts and getting appropriate responses based on simulated scenarios.
3. Compile a list of at least five potential use cases and associated constraints for implementing ChatGPT in upcoming medical scenarios.

Workshop Design – The study was designed around a single-session workshop lasting one hour. The session was divided into a 30-minute learning session and a 30-minute active hands-on exercise utilizing ChatGPT (Appendix A, available in online version). The learning session was designed to provide students with an understanding of ChatGPT, including its functionality, potential applications, and limitations. The learning session began with a discussion of participant familiarity, thoughts, and rumors about ChatGPT. We then transitioned to a 15-minute presentation about how ChatGPT is made and how it works using an analogy of opening a bakery (more details to be found in Appendix A). The second half of the workshop, the active half, concluded with an active demo that allowed students to interact with ChatGPT and discuss its potential uses in a medical setting. In this segment, students worked in pairs to go

through a case-based learning prompt from a small group session using ChatGPT. They also prompted ChatGPT to identify relevant medications and clinical guidelines, and to write board-style questions. We concluded with a 5-minute discussion on ChatGPT “hallucinations,” meaning instances where the model generates responses that are factually incorrect, fabricated, or not grounded in the input data or established knowledge, despite appearing coherent and plausible. We further discussed limitations with the tool and that it should be used as a learning aide rather than a thought and problem-solving replacement.

Surveys – We aimed to evaluate participants in four categories with respect to ChatGPT usage: frequency, use cases, comfort, and perceived future application. We collected attitudinal data at three time points: immediately before the workshop, immediately after the workshop, and then eight weeks post-workshop. Using a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree), we asked the same attitudinal questions in each survey about the participants’ use, and perceived future use, of ChatGPT, self-perceived understanding of how ChatGPT works at a basic level, and comfort using ChatGPT (Appendix B, available in online version). Additionally, participants were allowed to provide narrative comments about their ChatGPT

usage and other feedback and comments regarding the workshop.

Data Analysis – To provide an overview of the participant characteristics, descriptive statistics were calculated for limited demographic variables such as age and program of study. Paired t-tests were utilized to compare the responses to the same eight-question surveys administered before and immediately after the workshop. This statistical test was chosen due to the dependent nature of the samples (i.e., responses from the same participants at different time points) and the normal distribution of the data. Paired t-tests were also used to compare the mean responses from the pre-workshop survey to the survey administered six weeks post-workshop, assessing the longevity of the workshop's impact. We considered p values <0.05 to be statistically significant.

RESULTS AND DISCUSSION

The final study cohort consisted of 51 students who completed all three surveys. Eight students were lost to follow-up and not included in the study population. The persons lost to follow up were three PA students and five MD students. The final study cohort’s background and exposure to ChatGPT are provided in Table 1. Study participants had a mean age of 25.4 years ($SD = 5.1$), with representation from MD students (80%), PA students (18%), and genetic counseling students (2%).

Table 1. Sample population (N=51)

<i>Has used ChatGPT before workshop?</i>	<i>Yes (%)</i>	<i>No (%)</i>	<i>Total n (%)</i>
Genetic Counseling	0 (0)	1 (100)	1 (2)
MD Student	24 (59)	17 (41)	41 (80)
First-year	13	9	21
Second-year	11	8	19
PA Student	3 (33)	6 (67)	9 (18)
First-year	3	4	7
Second-year	0	2	2
Total n (%)	27 (53)	24 (47)	51

At baseline, slightly more than 50% of study participants reported ChatGPT usage, and less than 20% overall reported usage of at least one day per week (Figure 1A). This increased to more than 60% at eight weeks post-workshop. At this same later timepoint, the most commonly reported uses for ChatGPT among the choices provided were: explanations of pathophysiology/disease mechanism (63%, n=32); looking up medical terminology (47%, n=24); and non-medical questions (45%, n=23) (Figure 1B). With this single one-hour introduction to ChatGPT, average comfort with its use rose dramatically from 2.2 on the 5-point Likert scale to 4.3 immediately after the workshop (Figure 1C, $p < .05$ compared to baseline).

The effect of training persisted at eight weeks with an average score of 4.3. While

there was an initial hesitancy, there was a noticeable increase in the number of students who reported they could teach someone else to use ChatGPT following the workshop (Figure 1D). Before the workshop, the average response to “I see myself using ChatGPT in medical school” was 3.2 on the 5-point Likert scale, and this rose to 4 immediately after the workshop ($p < .05$ compared to baseline) and persisted at eight weeks (Likert scale 4.4, $p < .05$ compared to baseline) (Figure 1F). Furthermore, participants reported a significantly higher likelihood of using ChatGPT in their future medical practice after attending the workshop, again with that effect persisting at eight weeks post-workshop (Figure 1E).

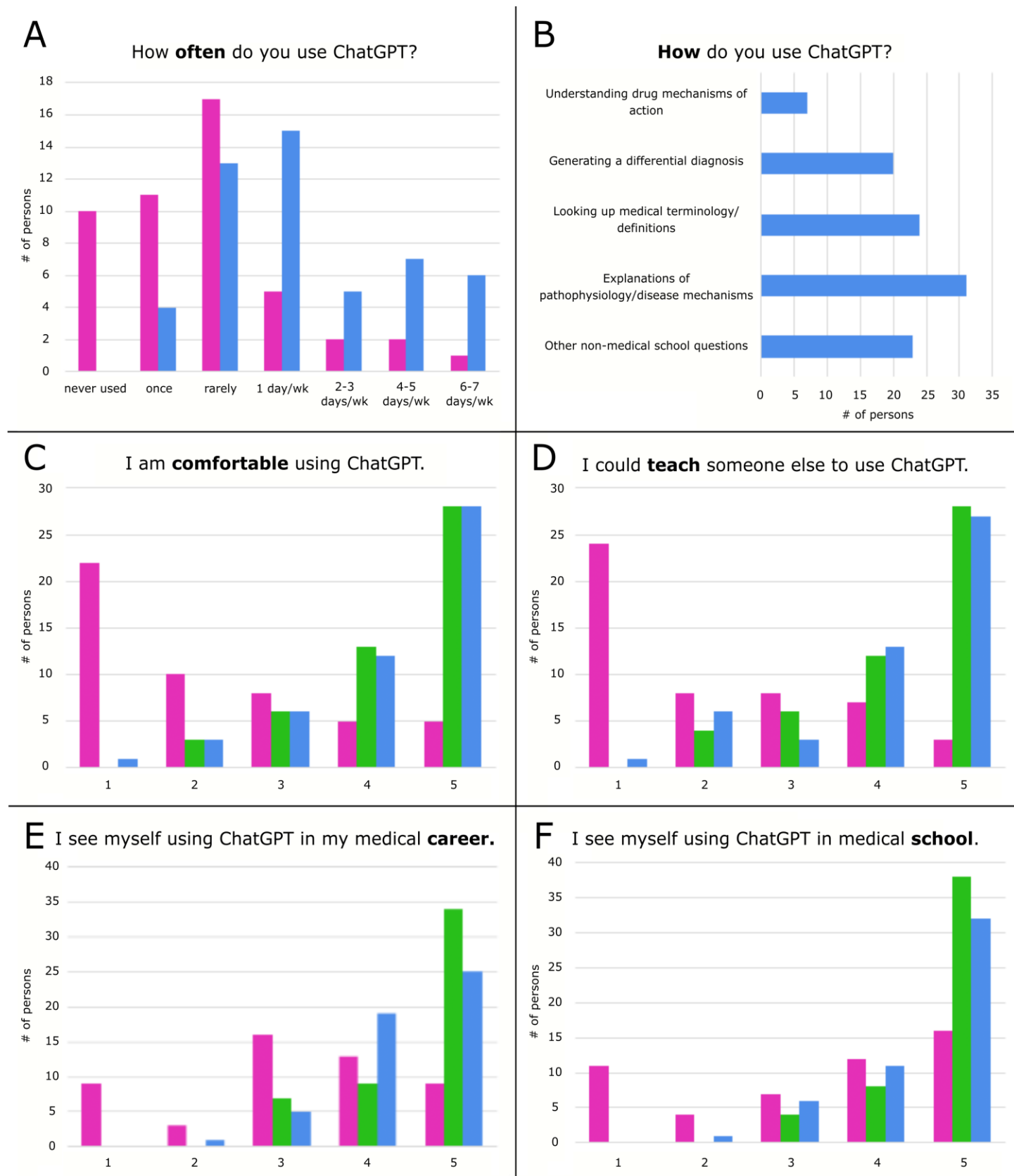


Figure 1. Survey responses at timepoints t1 (pink, pre-workshop), t2 (green, immediately post-workshop), and t3 (blue, eight weeks post-workshop) A. Frequency of ChatGPT use, B. Applications of ChatGPT reported at t3, C-F Respondent attitudes on 5-point Likert scale, ranging from 1, "Strongly Disagree" to 5, "Strongly Agree"

This initial study demonstrates the effectiveness of education and training in improving a group of medical, physician assistant and genetic counseling students' understanding and potential longer-term use of ChatGPT. The significant improvement in comfort level and likelihood of future use suggests that such workshops – and subsequent curricular integration – could be a valuable tool in integrating AI technologies like ChatGPT into health professionals' education and practice. The study also highlights the potential of this type of tool-based workshop for interprofessional training because it provides a common experience through which students from different disciplines can interact and learn. This may assist in 'de-siloing' health professions education and promote a more holistic approach to healthcare training.

CONCLUSIONS

Limitations

The study cohort consisted of self-selected students who chose to participate in the workshop out of personal interest or perceived relevance to their future medical careers. This could introduce a bias as these students might already have a positive inclination towards the use of AI in medical education. Moreover, the participants may have been influenced by increased

popularity of ChatGPT in the media and society-at-large. The follow-up period was limited, at eight weeks post-workshop. This duration might not be sufficient to assess the long-term impact of the workshop on the students' understanding and use of ChatGPT. As far as understanding, the study relied on self-reported measures of understanding and use of ChatGPT, which might not accurately reflect the actual understanding and use of the technology. Furthermore, the study lacked an assessment of students' ability to discern inaccuracies in ChatGPT responses.

Future Directions – Future research could involve longitudinal studies to assess the long-term impact of such workshops on the understanding, use, and perceptions of ChatGPT in medical education. Assessing usage amongst more diverse populations of students from different institutions, regions, and disciplines would similarly enhance the generalizability of the findings. Our study implemented scales for assessment; however, future studies may incorporate objective measures of understanding and use of AI, such as performance on tasks or assignments using ChatGPT and similar LLMs. For students participating in clinical care, the reliance of ChatGPT in patient care settings could be assessed – such as accuracy of diagnosis, treatment plans, and patient satisfaction. A future workshop on the evaluation of ChatGPT

responses, particularly guidance on discerning inaccuracies in these responses would be useful. This could be followed by an assessment of the students' ability to identify inaccuracies, considering the known capacity for ChatGPT to "hallucinate," as discussed earlier.

Ethical Considerations – As AI becomes more integrated into medical education and practice, future research should also explore the ethical considerations and potential pitfalls of such technology in academic medicine. This could include issues of data privacy, patient autonomy, academic integrity, and reliance on AI for decision-making.

Implications for Medical Education – LLMs are likely to transform the future of medical education. Tools such as ChatGPT provide students with instant access to a vast amount of medical knowledge, making them valuable for self-directed learning. Furthermore, LLMs

can lead to more personalized education by adapting to the individual learning needs and styles of each student – be it personalized assessments for board exams, assistance in determining a differential diagnosis, or guidance in scientific writing. Students can also be encouraged to critically evaluate the medical information that ChatGPT provides almost instantaneously, thereby enhancing critical thinking and decision-making skills. The integration of AI technologies like ChatGPT in health professions education also provides an opportunity to train students on the ethical aspects of using such technologies. Lastly, the use of ChatGPT supports the concept of continual learning, which is increasingly important in the rapidly evolving field of medicine, especially as AI and LLMs become increasingly accurate. However, further research is needed to fully understand the implications and potential of this technology in medical education.

ACKNOWLEDGMENTS

The authors received no grant support for this work. The authors declare that they received no payment or services from any third party to support this work.

REFERENCES

- Alam, F., Lim, M. A., & Zulkipli, I. N. (2023). Integrating AI in medical education: Embracing ethical usage and critical understanding. *Frontiers in Medicine*, 10, 1279707. <https://doi.org/10.3389/fmed.2023.1279707>
- Duong, D., & Solomon, B. D. (2023). Analysis of large-language model versus human performance for genetics questions. *European Journal of Human Genetics*, 32, 466–468. <https://doi.org/10.1038/s41431-023-01396-8>
- Eriksen, A. V., Möller, S., & Ryg, J. (2023). Use of ChatGPT-4 to Diagnose Complex Clinical Cases. *NEJM AI*, 1(1). <https://doi.org/10.1056/AIp2300031>
- Hirosawa, T., Kawamura, R., Harada, Y., Mizuta, K., Tokumasu, K., Kaji, Y., . . . Shimizu, T. (2023). ChatGPT-Generated Differential Diagnosis Lists for Complex Case-Derived Clinical Vignettes: Diagnostic Accuracy Evaluation. *JMIR Med Inform*, 11, e48808. <https://doi.org/10.2196/48808>
- Jeyaraman, M., K, S. P., Jeyaraman, N., Nallakumarasamy, A., Yadav, S., & Bondili, S. K. (2023). ChatGPT in Medical Education and Research: A Boon or a Bane? *Cureus*, 15(8), e44316. <https://doi.org/10.7759/cureus.44316>
- Kanjee, Z., Crowe, B., & Rodman, A. (2023). Accuracy of a Generative Artificial Intelligence Model in a Complex Diagnostic Challenge. *JAMA*, 330(1), 78–80. <https://doi.org/10.1001/jama.2023.8288>
- Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*, 2(2), e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
- Mehnen, L., Gruarin, S., Vasileva, M., & Knapp, B. (2023). ChatGPT as a medical doctor? A diagnostic accuracy study on common and rare diseases. *medRxiv*, 2023.2004.2020.23288859. <https://doi.org/10.1101/2023.04.20.23288859>

Mohammad, B., Supti, T., Alzubaidi, M., Shah, H., Alam, T., Shah, Z., & Househ, M. (2023). The Pros and Cons of Using ChatGPT in Medical Education: A Scoping Review. In J. Mantas, P. Gallos, E. Zoulias, A. Hasman, M. S. Househ, M. Charalampidou, & A. Magdalinou (Eds.), *Studies in Health Technology and Informatics*. IOS Press.
<https://doi.org/10.3233/SHTI230580>

Palmer, K. (2023, July 18). 'Doctors need to get on top of this': ChatGPT-4 displays bias in medical tasks. *STAT*. <https://www.statnews.com/2023/07/18/gpt4-health-disease-diagnosis-treatment-ai/>

Roose, K. (2023, January 12). Don't Ban ChatGPT in Schools. Teach With It. *The New York Times*. <https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html>

Shivaprakash, A. (2023, February 18). The Rise of LLM (Large Language Model) Chatbots. *Medium*. <https://medium.com/@arunashivaprakash575/the-rise-of-llm-large-language-model-chatbots-597009dd0724>

Shorey, S., Mattar, C., Pereira, T. L.-B., & Choolani, M. (2024). A scoping review of ChatGPT's role in healthcare education and research. *Nurse Education Today*, 135, 106121.
<https://doi.org/10.1016/j.nedt.2024.106121>

Subbaraman, N. (2023, April 28). ChatGPT Will See You Now: Doctors Using AI to Answer Patient Questions. *The Wall Street Journal*. <https://www.wsj.com/articles/dr-chatgpt-physicians-are-sending-patients-advice-using-ai-945cf60b>

Tan, S., Xin, X., & Wu, D. (2024). ChatGPT in medicine: prospects and challenges: a review article. *International Journal of Surgery*, 110(6), 3701-3706.
<https://doi.org/10.1097/js9.0000000000001312>

Tsang, R. (2023). Practical Applications of ChatGPT in Undergraduate Medical Education. *Journal of Medical Education and Curricular Development*, 10, 238212052311784.
<https://doi.org/10.1177/23821205231178449>

AUTHOR NOTES

Gabrielle Gershon

Assistant Research Staff, Department of Radiology, Medical Student, Emory University School of Medicine, Emory University

David Kulp

Research Fellow, Department of Medicine; Medical Student, Emory University School of Medicine, Emory University

Natalie Skopicki

Medical Student, Emory University School of Medicine, Emory University

Marly van Assen, PhD,

Assistant Professor, Department of Radiology, Emory University School of Medicine, Emory University

Judy Gichoya, MD

Associate Professor, Department of Radiology, Emory School of Medicine, Emory University