Erratum

DOI: 10.4274/ThoracResPract.2025.2025-6-3

Karakoyun ÖF, Koyuncuoğlu HE, Sağnıç ÖH, Özdemir ME, Gölcük Y, Yıldırım B. AI in patient care: evaluating large language model performance against evidence-based guidelines for pulmonary embolism. *Thorac Res Pract.* 2026;27(1):38-46.

The errors were made inadvertently by the authors.

i. The P value is incorrect in the "Results" subsection of the Abstract on page 38.

No significant difference was found in total scores **(*P* = 0.390)**.

The *P* value has been corrected in the "Results" subsection of the Abstract on page 38.

No significant difference was found in total scores **(*P* = 0.212)**.

ii. Information regarding the statistical test used for the overall comparison is missing in the "Statistical Analysis" section on page 42.

The following sentence has been added to the "Statistical Analysis" section on page 42:

Overall differences among the four LLM-based chatbots across the ten case-based questions were evaluated using the Friedman test, considering each question as a repeated-measures block.

iii. The statistical test, test statistic, and P value are incorrect in the "Results" section on page 42.

Although ChatGPT-4o obtained the highest total score and DeepSeek-V2 obtained the lowest, no statistically significant differences were observed among the AI models in overall performance ($H = 3.013$, $P = 0.390$).

The statistical test, test statistic, and *P* value have been corrected in the "Results" section on page 42.

Although ChatGPT-4o obtained the highest total score and DeepSeek-V2 the lowest, the Friedman test showed no statistically significant difference among the four LLM-based chatbots in overall performance across the ten questions, $\chi^2 = 4.50$, $P = 0.212$.

The relevant sections of the article have been corrected accordingly.