



ECG Interpretation: Humans Vs Machines

Take Home Point: In this study, cardiologists and emergency physicians (EP) had greater accuracy for electrocardiogram (ECG) interpretation than current versions of ChatGPT and Gemini.

Citation: Günay S, Öztürk A, Yiğit Y. The accuracy of Gemini, GPT-4, and GPT-4o in ECG analysis: A comparison with cardiologists and emergency medicine specialists. *Am J Emerg Med.* 2024 Oct;84:68-73. doi: 10.1016/j.ajem.2024.07.043.

Relevance: Artificial Intelligence (AI) is rapidly evolving and becoming an integral part of clinical medicine and research with large-language models (LLM) being evaluated for many novel applications traditionally performed by computers.

Study Summary: This study evaluated the performance of GPT-4, GPT-4o, and Gemini LLMs in ECG interpretation compared their accuracy to cardiologists and emergency medicine (EM) specialists. The authors of the study selected 20 routine ECGs and 20 “challenging” ECGs from a book of ECG cases (*150 ECG Cases*). The ECGs were presented with multiple choice responses and given to 12 cardiologists and 12 EM specialists to evaluate. The same questions were entered into GPT-4, GPT-4o, and Gemini Advanced on separate chat interfaces for review.

Importantly, the authors found there was no statistically significant difference between cardiologists and EM specialists in routine ECG interpretation or more challenging ECG questions. Cardiologists performed better than GPT-4, GPT-4o, and Gemini Advanced in routine ECG questions and the more challenging ECG questions. EM specialists outperformed the LLMs in routine ECG and overall ECG question, but not “challenging” ECGs specifically.

Among the LLMs, GPT-4o was the most accurate for ECG interpretation, especially among the more challenging cases.



Ivan Koay MBChB, MRCS, FRNZCUC, MD, is an Urgent Care Physician and Medical Lead for Kings College Hospital Urgent Treatment Centre, London, United Kingdom. He is also the Convenor for the Ireland and UK Faculty of the Royal New Zealand College of Urgent Care.

Editor’s Comments: The LLMs evaluated were not developed or specifically trained for ECG interpretation. It is unclear how more specifically trained algorithms may have compared to others and human experts. The study design also used multiple choice questions, which does not mimic real-world clinical practice. Time taken to interpret ECGs was not measured and change in performance based on time of day and ECGs were read without the normal distractions clinicians face in practice settings. For these reasons, these results should be taken with a grain of salt. Future studies should focus on comparing task specific LLMs to clinicians in more realistic settings since this is ultimately the most relevant question for what their role might be in patient care. ■

Laughter is the Best Medicine for Dry Eyes

Take Home Point: Laughter exercises were found to be non-inferior to 0.1% sodium hyaluronic acid in relieving symptoms of patients with dry eye disease.

Citation: Li J, Liao Y, Zhang S, et. al. Effect of laughter exercise versus 0.1% sodium hyaluronic acid on ocular surface discomfort in dry eye disease: non-inferiority randomized controlled trial. *BMJ.* 2024 Sep 11;386: e080474. doi: 10.1136/bmj-2024-080474

Relevance: Dry eye disease is a common and frustrating chronic condition. While many possible treatments exist, 0.1% sodium hyaluronic acid is among the most widely used forms of artificial tears and has been to relieve ocular discomfort. Prior studies have suggested mechanisms whereby laughter may increase tear production, including through the action of oxytocin.

Study Summary: This was a 2 arm, non-inferiority randomized controlled trial (RCT) in a tertiary care center in Southern China. Participants with symptomatic dry eye disease were recruited and block randomized in a 1:1 ratio to receive either “laughter exercise” or 0.1% sodium hyaluronic acid for 8 weeks. The laughter exercise group was instructed to perform the laughter exercise 4 times daily. The exercises required participants to repeat the phrases: “Hee hee hee, hah hah hah, cheese cheese cheese, cheek cheek cheek, hah hah hah hah hah hah.” They completed 30 rounds each time, lasting for at least 5 minutes (videos

“There’s also no reason not to recommend adding this to other therapies for dry eyes as there may be additional benefits for mood and other aspects of health.”

of the exercise are included in online version of the *BMJ* article). The control group applied artificial tears, 0.1% sodium hyaluronic acid eyedrops, to both eyes 4 times daily for 8 weeks.

The authors included and randomized 299 adult participants. The average age was 29 years old and 74% of the subjects were women. They found improvement in ocular surface disease index scores from baseline (after a washout period) to 8 weeks in both groups –10.5 points (95% CI –13.1 to –7.82) in the laughter exercise group and –8.83 (–11.7 to –6.02) in the control group (i.e., hyaluronic acid drops). These changes were statistically significant for both groups compared to baseline measurements after the 2-week washout period. Additionally, they found that laughter exercise improved tear film stability and the meibomian gland function.

Editor’s Comments: Blinding was not possible in this study given that participants needed to either perform the exercises or use a pharmaceutical agent. There may be limited generalizability due to the small sample size and single center setting in an ethnically homogenous location.

The upshot of this study is its use of a non-pharmacologic intervention in a well-designed RCT. Such studies are relatively rare given difficulty securing funding for behavioral interventions. Additionally, laughter has limited side effects (although it may not always be practical, especially in public, to perform these exercises). There’s also no reason not to recommend adding this to other therapies for dry eyes as there may be additional benefits for mood and other aspects of health. ■

Arm Position in Blood Pressure Measurements: Does it Really Matter?

Take Home Point: This study showed that common mis-

placements of the arm during blood pressure (BP) measurement (ie, on the lap or by the side) somewhat overestimates blood pressure, which may result in misdiagnosis of hypertension or erroneously missing hypotension. The differences were small on average but could be clinically significant in certain cases.

Citation: Liu H, Zhao D, Sabit A, et. al. Arm Position and Blood Pressure Readings: The ARMS Crossover Randomized Clinical Trial. *JAMA Intern Med.* 2024 Oct 7: e245213. doi: 10.1001/jamainternmed.2024.5213.

Relevance: BP measurement, like all objective data collection, requires standardized measurement approaches to ensure precise and accurate BP measurements. Guidelines for BP measurement include selecting the appropriate cuff size, cuff position, and measurement with the arm supported on a desk or table at the level of the heart level. Appropriateness of arm position is commonly overlooked when performing BP measurements.

Study Summary: This was a randomized crossover trial conducted among adults in Baltimore, Maryland. Each subject had measurements taken in 3 positions in random order: arm supported on a desk with mid-cuff at approximately mid-heart level (desk 1); hand supported on the lap (lap); and arm hanging at the side (side). All participants underwent a 4th set of triplicate BP measurements with the arm supported on a desk with mid-cuff at mid-heart level (desk 2), which is the same condition as desk 1 to account for any variability. Participants were recruited via BP screening program at a public food market, direct mail to previous study participants and information brochure placed in hypertension clinics at Johns Hopkins University. Measurements took place during daytime clinic hours using a validated oscillometric BP device (ProBP 2000 Digital Blood Pressure Device [Welch Allyn])

The authors enrolled 133 participants. They found average BP measurements were: 126/74 mm Hg for each of the desk 1 and desk 2 positions; 130/78mm Hg for the lap position; and 133/78mm Hg for the side position with results consistent across subgroups. The lap position overestimated systolic blood pressure (SBP) and diastolic blood pressure (DBP) by approximately 4 mmHg, whereas the side position overestimated SBP by 6 mmHg and DBP by approximately 4 mmHg.

Editor’s Comments: This study was limited to a single urban center and had relatively small numbers in each group. Additionally, they examined only a single, automatic BP cuff. It’s unclear to what extent these trends may have

been observed with manual BP measurement or other automatic devices. Regardless, this study does highlight the importance of standardized arm positioning to ensure BP measurements are recorded accurately and the values are comparable between occasions and locations. ■

Factors Driving Increased Pediatric Urgent Care Demand

Take Home Point: A combination of declining primary care access, circulating viral infections, and changing patterns of chief complaints were associated with increases in frequency and duration of visits to pediatric UC.

Citation: Lehan E, Briand P, O'Brien E, et. al. Synergistic patient factors are driving recent increased pediatric urgent care demand. *PLOS Digit Health*. 3(8): e0000572

Relevance: As UC services evolve, understanding factors that drive presentations and affect patient volumes is necessary for UC administrators and clinicians to be best prepared for surges in patients presenting to pediatric UC centers. Such data will also help UC to choose appropriate levels of staffing.

Study Summary: This was a retrospective cohort study reviewing a local healthcare center's National Ambulatory Care Reporting System data in Canada, with data collected from their electronic health record database. The authors looked to use high-fidelity NACRS data to model and identify factors contributing to the increased demand for pediatric urgent care (PUC). Data included aggregate PUC visits from April 2006 to December 2022 in this region of Canada.

The authors retrospectively analyzed a total of 164,660 visits during the study period. There was an increase in the number of visits per day on average, with daily volumes increasing in 2015. This trend abated in 2020 at the start of the COVID-19 pandemic and then rapidly resumed to previous levels in 2021 and 2022. The authors found an increase in the absolute numbers of all levels of acuity across the study period. The trend is most notable for "urgent" level presentations with more than triple the urgent presentations in 2022 when compared to 2007. Patients without identified primary care clinicians were more likely to present with both "emergent" level and "non-urgent" level presentations and were also more often diagnosed with mental health conditions. The authors also noted that increased levels of circulating infectious diagnoses, and shifts in chief complaints were driving increased frequency and duration of visits.

Editor's Comments: This study lacks granular data analysis given its retrospective design and use of aggregated data. There also may be limited generalizability as the PUC center data was obtained from a medium size city in Canada. The study does suggest that PUC in particular relies on an understanding of infectious disease epidemiology and primary care access to best predict patient volumes and needs. Emergency medicine organizations have cited the need for primary care access for years as a strategy to mitigate overcrowding and improve healthcare outcomes for patients. This study provides data suggesting UC is susceptible to similar crowding and decreased efficiency when access to primary care for children is insufficient. ■

What Would Happen if We Didn't Treat Strep Throat?

Take Home Point: In this small Swiss RCT, treatment for group A streptococcal (GAS) pharyngitis in children with placebo was non-inferior to treatment with amoxicillin for reducing fever duration, while pain intensity and risk of complications were similar in this study.

Citation: Gualtieri R, Verolet C, Mardegan C, et. al. Amoxicillin vs. placebo to reduce symptoms in children with group A streptococcal pharyngitis: a randomized, multi-center, double-blind, non-inferiority trial. *Eur J Pediatr*. 2024 Nov;183(11):4773-4782. doi: 10.1007/s00431-024-05705-1.

Relevance: Pharyngitis is among the most common UC complaints for both children and adults. Current guidelines suggest antibiotics should be used almost exclusively for GAS pharyngitis. Over recent decades, however, controversy has emerged about the necessity and benefit of antibiotic treatment, even in confirmed cases of GAS pharyngitis.

Placebo-controlled trials require that there is reasonable belief in clinical equipoise between treatment and non-treatment of a certain condition. As growing evidence emerges over the various risks of antibiotics, doubt also

“Current guidelines suggest antibiotics should be used almost exclusively for GAS pharyngitis.”

has grown over the benefit of antibiotics in GAS pharyngitis treatment for symptom improvement and prevention of suppurative and non-suppurative complications. This study aims to sort out what benefit (if any) exists for prescribing antibiotic therapy for children with GAS pharyngitis in high-income countries.

Study Summary: This was a prospective, double-blind, randomized, non-inferiority clinical trial in children with GAS pharyngitis who presented to the emergency departments of 2 pediatric university hospitals and 1 regional hospital in Switzerland. Children were randomly assigned in a 1:1 ratio to receive either a 6-day placebo regimen (intervention group) or amoxicillin tablets (control group). Randomization was stratified by weight groups (<18 kg, 18–24 kg, and >24 kg) and study centers. Amoxicillin dosing was weight-based to achieve the recommended dose of 50 mg/kg/day divided in 2 doses (BID). All patients had a rapid strep test and culture to confirm the diagnosis of GAS. The primary outcome was the difference in the fever duration with the threshold for non-inferiority of 12 hours difference. Secondary outcomes included pain intensity, use of analgesics, treatment failure (defined as any complication or clinical deterioration that the treating clinician felt warranted starting an antibiotic), persistence of symptoms on day 3, GAS pharyngitis complications, and GAS eradication rate at 1 month after enrollment.

The authors recruited and randomized 88 children. However, only 65 children (31 in amoxicillin group and 35 in treatment group) adhered to the treatment schedule. The mean duration of fever was 21.7 hours in the amoxicillin group and 24.6 hours in the placebo group with a mean difference in fever duration of only 2.8 h (95% CI, -6.5 to 12.2). The authors observed no statistically significant difference between the 2 groups regarding the use of symptomatic treatment (acetaminophen or non-steroidal anti-inflammatory drugs, or NSAIDs) at day 3.

Treatment failure occurred in 13% (6 patients) in the placebo group and 5% (3 patients) in the antibiotics group. “Treatment failure” in both groups was most commonly otitis media or scarlet fever. One case of retropharyngeal abscess was diagnosed in the placebo group. The relative risk (RR) of treatment failure, therefore, was 2.15, but was not statistically significant (95% CI, 0.44 to 10.57). No subsequent suppurative or non-suppurative complications were observed in either group in the 12-month follow-up period.

There was a significantly higher rate of persistently positive throat culture in the placebo group compared to the amoxicillin group at 30-days after the initial visit, (67% vs. 15%, $p=0.002$; $RR = 4.44$ for positive culture at one month.

Editor’s Comments: While this was a very small study, it was well designed. Unfortunately, there were issues enrolling more patients related to parental consent and acceptance as well as factors that arose related to the COVID-19 pandemic. Regardless, this study is a landmark study of sorts and will likely garner attention for years to come. Given concerns for complications from untreated GAS pharyngitis, it has been deemed unethical for decades to randomize patients to not receive antibiotics. However, based on more recent data, definitive benefit of using antibiotics for treating this condition has become less clear, especially in the developed world. These results should be interpreted with extreme caution by clinicians in lower income countries and areas with rates of rheumatic fever as these results are likely not generalizable to these settings. Notwithstanding, this study does open the door for further study and even shared-decision making within UC for patients and parents whose values favor medication avoidance. Additionally, despite its small size, we can glean some insights from this small pediatrics study: in the developed world, antibiotics don’t seem to do much to shorten the course of illness or reduce symptoms, and children who are not treated with antibiotics are more likely to continue to have GAS present in the oropharynx. ■

Accuracy of Clinician Interpretation of Pediatric Elbow Radiographs

Take Home Point: In this non-clinical exercise, experienced healthcare professionals overcalled a diagnosis of significant injury nearly 50% of the time when examining pediatric elbow radiographs.

Citation: Dann L, Edwards S, Hall D, et al. Black And White: How Good Are Clinicians At Diagnosing Elbow Injuries From Pediatric Elbow Radiographs Alone? *Emerg Med J.* 2024; 41:662–667

“Among participants, 49.7% reported an injury on the normal radiograph.”

Relevance: Because of complex patterns of ossification, pediatric elbow x-rays (XR) are difficult to assess for many

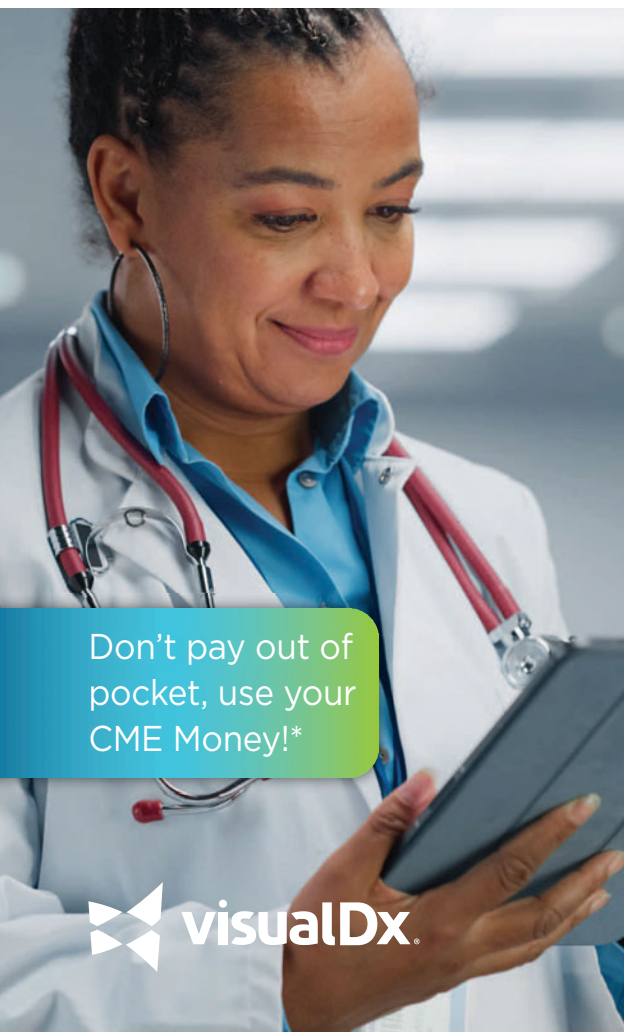
non-radiologists. Missing or overcalling fractures can have negative implications on patient outcomes and resource utilization.

Study Summary: This prospective study was conducted via the Free Open Access Medical Education (FOAMed) platform, Don't Forget the Bubbles (DFTB, ISSN 2754-5407). It consisted of 2 parts: a participant survey and a reporting exercise. The survey consisted of 9 questions including participant demographics, specialty, years of postgraduate clinical experience and experience with pediatric elbow XR interpretation. Clinician participants were asked to rate their proficiency in interpreting trauma XRs. The exercise consisted of participants reviewing 10 trauma XRs obtained from a single tertiary pediatric ED within a 20-minute period.

Among those recruited, 318 clinicians (76%) reported that they routinely interpret pediatric elbow XRs in their current clinical role. These 318 clinicians were included in the analysis. The participants comprised 237 physicians, and the remainder were advanced practice providers, nurses, and unidentified healthcare professionals. The clinicians generally had considerable experience with 72.3% of participants having >6 years of clinical experience.

Completely accurate interpretation of all 10 XRs was rare with only 9/318 (2.8%) of correctly identifying whether a fracture was present or not in all images. The Gartland 3 supracondylar fracture was reported correctly most frequently; the lateral condyle fracture, conversely, was reported incorrectly most frequently. Among participants, 49.7% reported an injury on the normal radiograph. The mean number of radiographs correctly interpreted was 5.44 but higher (6.02) for those with >6 years of experience. Emergency Medicine (EM) and Pediatric Emergency Medicine (PEM) clinicians had similar accuracy and were both more accurate than general practitioners (PEM).

Editor's Comments: This was an artificial study outside of a real-world clinical practice setting. The clinicians were not able to obtain a history or examine the patients. It is difficult to ascertain what these results suggest about the actual abilities of these clinicians in interpreting pediatric elbow XR in practice. The rate of fractures being overcalled was high but fits with the clinical reality that EM and UC clinicians commonly will err on the side of conservatism (ie, immobilizing a possible fracture), especially in the care of injured children. ■



Don't pay out of pocket, use your CME Money!*



Solve Challenging Cases Quickly

How VisualDx Can Help:

- Save time
- Improve operational efficiency
- Reduce unnecessary referrals

JUCM Readers Get
\$50 OFF
visualdx.com/jucm

*We accept CME reimbursement, as long as your employer allows it. Check with them to confirm. Learn more: visualdx.com/earn-cme