

A Retrospective Analysis of Ophthalmology Triage Accuracy: Comparing the Performance of Ophthalmologists, Specialist Nurses, Staff Nurses, and an AI Triage System

Jaskaran Singh Bhangu^{1*}, Mohamed Morgan², Rhadika Rewal¹, Christopher Stewart¹, Gwyn Williams¹, Safa Elhassan¹

¹Department of Ophthalmology, Singleton Hospital, Swansea, UK

²Nottingham University Hospitals, Nottingham, UK

Email: *2018939@swansea.ac.uk

How to cite this paper: Bhangu, J.S., Morgan, M., Rewal, R., Stewart, C., Williams, G. and Elhassan, S. (2025) A Retrospective Analysis of Ophthalmology Triage Accuracy: Comparing the Performance of Ophthalmologists, Specialist Nurses, Staff Nurses, and an AI Triage System. *International Journal of Clinical Medicine*, 16, 539-547.

<https://doi.org/10.4236/ijcm.2025.1612040>

Received: November 3, 2025

Accepted: December 6, 2025

Published: December 9, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Purpose: To evaluate and compare the efficiency and accuracy of telephone triage of unscheduled eye care in relation to their respective roles (ophthalmologists, specialist nurses, staff nurses, and an AI system). **Methods:** In this retrospective study, we evaluated 2581 unscheduled eye care telephone triage forms from Singleton Hospital between 2022 and 2024. The four main data cohorts were ophthalmologists (929 forms), specialist nurses (900 forms), staff nurses (500 forms), and an AI system (ChatGPT-4O), which evaluated 252 clinical details. The primary outcome measures were adherence to “traffic-light” guidelines, rates of case escalation and de-escalation, and final triage outcome (appointment/advice, etc.). **Results:** There was no statistical difference in the percentage of adherence to the “traffic-light” guidelines between the three human groups. Specialist and staff nurses were significantly more likely to escalate a case (7% and 8%, respectively, compared to 4% for ophthalmologists); however, this resulted in no additional out-of-hours work. Ophthalmologists had the highest rate of accurate diagnostics from provisional to actual diagnosis being 46%, and managed the largest number of triage calls. Both specialist and staff nurses stated they were stressed by the triage process and also that casualty doctors frequently interrupted them while triaging, much less so for ophthalmologists. The AI system showed some promise; however, it needed further refinement. **Conclusion:** Although the different staff roles employed for telephone triage produce comparable patient outcomes, there are significant differences in terms of case escalation, diagnostic accuracy, and the impact on the clinic’s workflow. The amount of stress both spe-

cialist and staff nurses report experiencing with respect to triage, and the frequency of interruption by casualty doctors, demonstrate a need to optimize the system. The use of AI as a tool for future risk stratification appears promising; however, it will require additional development.

Keywords

Ophthalmology, Urgent Eye Care, Artificial Intelligence

1. Introduction

Demand for urgent and emergent eye care globally is creating significant stressors for health-care systems worldwide. As illustrated in reports, urgent eye appointments are increasing annually by 3.8% - 7.9% [1] and are increasing exponentially. This particular trend is especially evident in the UK, where visits to eye emergency departments increased by 7.9% annually between 2010 and 2018. Projecting the current trend, there may be a doubling of attendance by 2025 [2]. The increasing volumes create a backlog of patients requiring specialist ophthalmology attention. Furthermore, the increasing volumes are compounded by the problem of inappropriate care. Approximately 30% - 37% of patients attending an eye emergency department could have been appropriately managed by a primary care provider [3]. Therefore, there exists a substantial need for effective front door triage processes to appropriately stratify patients and direct true emergencies to specialist ophthalmologists in a timely manner and appropriately divert non-emergencies to community or primary care providers.

Tele-triage is increasingly becoming the first point of contact for patients seeking unscheduled care. Tele-triage creates several challenges. Triage clinicians, typically nurses, are making high-risk, high-stakes decisions solely based on what a patient verbally describes as being wrong with them and are doing so remotely without having the opportunity to visually examine the patient or obtain any type of diagnostic imaging [3]. This remote assessment is further complicated by issues including a patient's difficulty articulating what is wrong with them, barriers to effective communication, and the pressure of a high volume of unseen work [4].

According to The Royal College of Ophthalmologists and The College of Optometrists, an urgent eye condition is "any condition that is of recent onset and is distressing or is believed by the patient, carer or referring health professional to present an imminent threat to vision or to general health." [4]. The triage clinician is often responsible for identifying those true threats and distinguishing them from benign conditions. While nurse-led triage is an accepted model [5]-[7], there is limited detailed data comparing the performance of different types of nursing personnel (*i.e.*, specialty vs. staff) to that of physicians (ophthalmologists), specifically in the context of telephone triage.

Additionally, the emergence of powerful large language models (LLM) provides a new and potentially useful option to assist or possibly even replace aspects of

triage. The use of Artificial Intelligence (AI) to support or replace triage tasks would represent a positive innovation for an overburdened system [8] [9].

The purpose of this research is to address this knowledge gap by comparing the effectiveness, accuracy, and broader implications of unscheduled, in-hours eye care telephone triage at a busy secondary care hospital. Our research uses a retrospective evaluation of triage performance, comparing the results of triage provided by ophthalmologists, specialist nurses who tend to have more experience and are able to apply more interventional techniques (Band 6 - 7), and staff nurses who are more senior than general nurses but not as advanced as specialist nurses (Band 5), and an initial assessment of an AI system (ChatGPT-4O).

2. Methods

This retrospective, comparative cohort evaluation took place in the Urgent Eye Care (URCE) service of Singleton Hospital, a busy secondary care ophthalmology department. All telephone triage forms submitted by clinicians and completed during business hours between 2022 and 2024 were extracted from the hospital's electronic patient record system. Forms containing incomplete information or relating to pre-scheduled follow-up appointments were excluded from the sample.

Using this exclusionary method, a total of 2581 eligible telephone triage sessions were identified. Sessions were categorized into four subgroups based on the professional completing the triage:

- Ophthalmologists: 929 triage sheets (sessions from 2023-2024).
- Specialist Nurses (Band 6-7): 900 triage sheets (sessions from 2022-2024).
- Staff Nurses (Band 5): 500 triage sheets (sessions from 2023-2024).
- AI (ChatGPT-4O): 252 unique clinical details from 2023 sessions were processed through the AI system.

Permission to conduct the study was granted by the Institutional Review Board for the location where the study took place. All identifiable information was removed from the data prior to analysis.

Measures and Data Analysis

The main outcome measures for comparison among the groups were: adherence to "traffic-light" guidelines, case management (escalation/de-escalation), and final triage outcome.

1) Guideline Compliance: Each session was assessed to determine if the clinician followed the locally established in house "traffic-light" triage guidelines. This protocol categorizes patients based on the self-reported signs and symptoms into the categories of Immediate (Red), *i.e.*, chemical injury, IOP > 40 mm Hg; Within 24 Hours (Amber), *i.e.*, acute anterior uveitis, IOP 30 - 40 mm Hg; and Standard/Non-urgent (Green), *i.e.*, corneal abrasion, IOP 24 - 30 mm Hg. A session was classified as compliant if the final triage outcome (e.g., "book immediate appointment") corresponded to the urgency level designated in the guidelines for the self-reported symptoms.

2) Case Management: Escalation was defined as a situation in which a triage clinician directed a patient to a higher level of care than the guideline indicated (e.g., booking an “Immediate” appointment for an “Amber” condition). De-escalation was defined as directing a patient to a lower level of care or managing with advice when the guideline indicated an in-person consultation.

3) Triage Outcome: The final outcome of each session was categorized into one of the five possible categories: (1) Appointment booked (with urgency indicated); (2) Self-management advised; (3) Prescription issued; (4) Referral to a General Practitioner (GP); or (5) Referral to hospital Accident & Emergency (A&E).

4) AI Performance: Anonymized clinical summaries of 252 triage forms were processed through ChatGPT-4O. Additionally, to further assess the AI’s diagnostic accuracy, a randomly selected set of 30 scenarios were tested. The AI’s recommended diagnosis and treatment plan were compared to the final, consultant-approved diagnosis in the patient record. Additional testing was done with 3 cases in which it was known that a human triage failed to adhere to the guidelines in order to determine the AI’s ability to correctly identify the appropriate pathway.

5) Qualitative Data: Short, anonymous surveys comprised of Likert scale and open-ended questions were sent to the triage nurses (n = 12) and casualty doctors who consist of ophthalmology registrars and consultants working in urgent rapid clinic for eyes (URCE) (n = 8) to gather perceptions regarding the degree of stress, workload, and interprofessional communication associated with the triage process.

Chi-square tests were used to assess significant differences in the proportion of adherence and escalation between groups. P-values less than 0.05 were considered statistically significant. Themes in qualitative comments were analysed.

3. Results

Although the results of the study reveal that there were similarities in the use of the “traffic light” guidelines amongst all of the groups of humans who used them, the study found that there was a statistically significant increase in the use of the guidelines amongst the senior nurses from 2022 through to 2024 ($p < 0.0001$). This increase in usage was directly related to the implementation of a new digital telephone triage template.

There were no statistically significant differences in the types of triage outcomes that occurred among the groups of humans. Similarly, although there were some minor differences in the non-appointment triage outcomes, all groups of humans made similar numbers of bookings for in-person appointments. The only group of humans that provided a greater number of “advice only” responses were the ophthalmologists, while the greatest number of prescriptions were written by the staff nurses. The prescriptions were typically for common conditions such as infectious conjunctivitis and blepharitis.

Statistically significant differences in case escalation occurred between the two groups of nurses and the ophthalmologists. The staff nurses and specialist nurses escalated approximately 8% and 7% of the cases, respectively, while the ophthal-

mologists escalated only approximately 4% of the cases. This was a statistically significant difference ($p < 0.01$). Although there was a statistically significant difference in the rate of escalation of cases, there were very few cases that required de-escalation. The rate of de-escalation was less than 2% and there were no statistically significant differences in the rate of de-escalation among the various groups of humans.

Although there were statistically significant differences in the rate of escalation of cases amongst the groups of humans, the increased rate of escalation of cases by the nurses did not result in an increased number of patients being sent to the out-of-hours services or the general Accident and Emergency department. The majority of the escalated cases were booked for in-hours appointments that were deemed to be “safe” for patients with borderline conditions or conditions that were suitable for advice.

Furthermore, when assessing the accuracy of the provisional vs actual diagnosis among the healthcare professionals, ophthalmologists had the highest accuracy with 46%, significantly higher than their colleagues specialist nurses and staff nurses who had accuracy rates of 27% and 24% respectively ($p < 0.05$).

The performance of the artificial intelligence (AI) was variable in the test of the 252 clinical details. The more focused test of 30 random clinical scenarios demonstrated that the AI had a diagnostic accuracy of 33.3% (10/30), based on whether the AI’s suggested diagnosis and management plan matched the final consultant-verified outcome.

Interestingly, 6 of the 10 correct cases were from scenarios that were originally triaged by an ophthalmologist. This suggests that the performance of the AI is heavily dependent upon the level of clarity, specificity, and clinical detail within the input data.

Additionally, a separate, targeted test of the AI was conducted with 3 cases in which the human triage was not compliant with the guidelines. The AI correctly identified the appropriate, compliant pathway in 2 of the 3 cases and requested additional information regarding the third case. This indicates a possible utility for the AI as either a quality assurance tool or a real-time decision support tool.

A significant theme that emerged from the qualitative survey data was that telephone triage was perceived as “stressful” or “very stressful” by all of the nurse respondents. The thematic analysis of the nurse respondents’ open-ended comments highlighted that there were three primary areas of concern: (a) the high volume of calls; (b) the uncertainty of making a clinical decision remotely (“not enjoyable when I don’t know what I’m doing”); and (c) the high-stakes nature of possibly failing to identify a sight-threatening condition [8].

In stark contrast, the casualty doctors perceived the URCE service as significantly less stressful when an ophthalmologist was the designated healthcare professional for triage. Additionally, the casualty doctors reported that when a nurse was the designated triage, they received frequent interruptions to their own clinics—approximately 4 - 6 per clinic session—to seek clarification, advice, or con-

firmation of a triage plan. Conversely, the casualty doctors reported that they rarely received interruptions to their clinics when an ophthalmologist was the designated triage.

This finding clearly indicates that while nurse-led triage achieves similar patient outcomes, it produces a “hidden workload” and imposes a significant cognitive burden on clinic-based medical staff, a well-documented factor contributing to clinical errors [9] [10]. Moreover, several of the doctors indicated that they would “frequently” or “almost always” conduct a full re-assessment of a case that was triaged by a nurse prior to providing treatment, a practice they felt was unnecessary when treating cases that were triaged by an ophthalmologist.

4. Discussion

Overall, the present study offers a unique and multifaceted comparison of telephone triage performance in an urgent eye care setting. The main finding of the present study was that, despite the fact that ophthalmologists, specialist nurses, and staff nurses achieved similar levels of compliance with the “traffic-light” guidelines and similar levels of final patient outcomes, there were many significant differences in terms of clinical practice, diagnostic accuracy, and system-wide efficiency.

The substantially higher rates of escalation of cases by nurses (8% and 7%, respectively) are likely due to the combination of the self-reported stress and clinical uncertainty [8] experienced by nurses in the present study, along with the existing literature documenting that clinicians experiencing clinical uncertainty or pressure will frequently engage in defensive practice, including lowering their risk thresholds [11]. In the context of the present study, this means that the nurse-led triage process includes an increased tendency to elevate borderline cases to an in-person appointment, thus producing a “safe” decision for both the patient and the practitioner. While this is not necessarily a negative finding—it does indicate a safety-first culture—it has clear implications for service efficiency, including transforming what could have been “advice-only” cases into appointments that occupy clinic capacity.

Perhaps the most important finding of the present study is the significant and heretofore unmeasured downstream effects of the triage process on the larger clinical workflow. The qualitative data collected from the doctors regarding the frequency of interruptions and the necessity to “re-triage” patients quantitatively document a substantial duplication of effort and a significant cause of clinical inefficiency [12] [13]. Thus, while the patient-facing outcomes of the nurse-led triage process appear to be equivalent to those produced by the ophthalmologist-led triage process, the “cost” in total staff time and cognitive load for the nurse-led triage process is significantly higher.

Furthermore, the stress reported by 100% of the nurses in the present study represents a critical finding in itself, indicating a service model that is creating unsustainable pressure on staff [14].

As anticipated, the ophthalmologists demonstrated superior diagnostic accuracy and lower escalation rates in comparison to the nurses. This is consistent with the expectation that the ophthalmologists, as specialists, have received more extensive and specialized training [15] [16]. Nevertheless, utilizing consultants or specialist registrars to answer all initial telephone triage calls is likely an expensive and unsustainable model for most public health systems. As such, the present study identifies the fundamental conflict in urgent care: the most diagnostically accurate triage is often the least cost-effective resource for the role.

While the performance of the AI system was variable in the test of 252 clinical details, the performance of the AI system was variable in the test of 252 clinical details. The more focused test of 30 random clinical scenarios demonstrated that the AI had a diagnostic accuracy of 33.3% (10/30), based on whether the AI's suggested diagnosis and management plan matched the final consultant-verified outcome.

Interestingly, 6 of the 10 correct cases were from scenarios that were originally triaged by an ophthalmologist. This suggests that the performance of the AI is heavily dependent upon the level of clarity, specificity, and clinical detail within the input data.

Additionally, a separate, targeted test of the AI was conducted with 3 cases in which the human triage was not compliant with the guidelines. The AI correctly identified the appropriate, compliant pathway in 2 of the 3 cases and requested additional information regarding the third case. This indicates a possible utility for the AI as either a quality assurance tool or as a real-time decision support tool.

A significant theme that emerged from the qualitative survey data was that telephone triage was perceived as “stressful” or “very stressful” by all of the nurse respondents. The thematic analysis of the nurse respondents’ open-ended comments highlighted that there were three primary areas of concern: (a) the high volume of calls; (b) the uncertainty of making a clinical decision remotely (“not enjoyable when I don’t know what I’m doing”); and (c) the high-stakes nature of possibly failing to identify a sight-threatening condition [8].

In stark contrast, the casualty doctors perceived the URCE service as significantly less stressful when an ophthalmologist was the designated triage. Additionally, the casualty doctors reported that when a nurse was the designated triage, they received frequent interruptions to their own clinics—approximately 4-6 per clinic session—to seek clarification, advice, or confirmation of a triage plan. Conversely, the casualty doctors reported that they rarely received interruptions to their clinics when an ophthalmologist was the designated triage. Nevertheless, it is also important to highlight the limitation of this study. Particularly in these qualitative findings being obtained from a relatively small pool (12 nurses, 8 doctors) and therefore its generalisability being weak.

This finding clearly indicates that, while nurse-led triage achieves similar patient outcomes, it produces a “hidden workload” and imposes a significant cognitive burden on clinic-based medical staff, a well-documented factor contributing

to clinical errors [9] [10]. Moreover, several of the doctors indicated that they would “frequently” or “almost always” conduct a full re-assessment of a case that was triaged by a nurse prior to providing treatment, a practice they felt was unnecessary when treating cases that were triaged by an ophthalmologist.

5. Conclusion

Unscheduled eye care telephone triage is a demanding and high-stakes role associated with significant stress for nursing staff and considerable workflow interruptions for clinic-based doctors. This study demonstrates that while ophthalmologists, specialist nurses, and staff nurses can all perform this triage to a similar standard of safety and patient outcome, this apparent equivalence is misleading. It masks crucial differences in diagnostic accuracy, case escalation, and system-wide efficiency. The high stress on nurses and the hidden workflow burden on doctors are not sustainable. To build a more resilient and efficient service, future efforts should focus on alternative, multidisciplinary staffing models, such as the integration of optometrist-led triage, and the continued development of AI tools to support, rather than replace, clinical decision-making.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Channa, R., Zafar, S.N., Canner, J.K., Haring, R.S., Schneider, E.B. and Friedman, D.S. (2016) Epidemiology of Eye-Related Emergency Department Visits. *JAMA Ophthalmology*, **134**, 312-319. <https://doi.org/10.1001/jamaophthalmol.2015.5778>
- [2] De Silva, S.R., *et al.* (2020) Patterns of Attendances to the Hospital Emergency Eye Care Service: A Multicentre Study in England. *BMJ Open Ophthalmology*, **5**, e000432.
- [3] Hau, S., Ioannidis, A., Masaoutis, P. and Verma, S. (2008) Patterns of Ophthalmological Complaints Presenting to a Dedicated Ophthalmic Accident & Emergency Department: Inappropriate Use and Patients’ Perspective. *Emergency Medicine Journal*, **25**, 740-744. <https://doi.org/10.1136/emj.2007.057604>
- [4] Asgari, E., *et al.* (2022) Exploring the Challenges to Telephone Triage in Pre-Hospital Emergency Care: A Qualitative Content Analysis. *BMC Nursing*, **21**, Article No. 253.
- [5] Greenberg, M.E., *et al.* (2010) Nurses’ Experiences with Telephone Triage and Advice: A Meta-Ethnography. *Journal of Advanced Nursing*, **66**, 2407-2418.
- [6] Chen, J., Chen, C., Zheng, Y. and Zhong, L. (2023) Corrigendum: Characteristics of Eye-Related Emergency Visits and Triage Differences by Nurses and Ophthalmologists: Perspective from a Single Eye Center in Southern China. *Frontiers in Medicine*, **10**, Article ID: 1091128. <https://doi.org/10.3389/fmed.2023.1241114>
- [7] Hau, S., Ehrlich, D., Binstead, K. and Verma, S. (2006) An Evaluation of Optometrists’ Ability to Correctly Identify and Manage Patients with Ocular Disease in the Accident and Emergency Department of an Eye Hospital. *British Journal of Ophthalmology*, **91**, 437-440. <https://doi.org/10.1136/bjo.2006.105593>
- [8] Vester, C., *et al.* (2024). Integrating Nurse-Led Interventions in Ophthalmology Care: A Systematic Review. *Journal of Clinical Medicine*, **13**, 3749.
- [9] Ting, D.S.W., Pasquale, L.R., Peng, L., Campbell, J.P., Lee, A.Y., Raman, R., *et al.*

- (2018) Artificial Intelligence and Deep Learning in Ophthalmology. *British Journal of Ophthalmology*, **103**, 167-175. <https://doi.org/10.1136/bjophthalmol-2018-313173>
- [10] Al-Ghamdi, S., et al. (2022). Artificial Intelligence (AI) in Ophthalmology: An Overview. *Ophthalmology Research: An International Journal*, **15**, 1-10.
- [11] Westbrook, J.I., Coiera, E., Dunsmuir, W.T.M., Brown, B.M., Kelk, N., Paoloni, R., et al. (2010) The Impact of Interruptions on Clinical Task Completion. *Quality and Safety in Health Care*, **19**, 284-289. <https://doi.org/10.1136/qshc.2009.039255>
- [12] Wu, A.M., et al. (2017) Accuracy of Referral Diagnosis to an Emergency Eye Clinic. *Canadian Journal of Ophthalmology*, **52**, 143-148.
- [13] O'Brien, B., et al. (2015) Defensive Practice in Telephone Triage: A Qualitative Study. *Nursing Standard*, **29**, 37-43.
- [14] Baker, H., et al. (2016) Evaluation of a Minor Eye Conditions Scheme Delivered by Community Optometrists. *BMJ Open Ophthalmology*, **1**, e000004.
- [15] The College of Optometrists and The Royal College of Ophthalmologists (2013) Commissioning Better Eye Care. Clinical Commissioning Guidance.
- [16] Westbrook, J.I., Raban, M.Z., Walter, S.R. and Douglas, H. (2018) Task Errors by Emergency Physicians Are Associated with Interruptions, Multitasking, Fatigue and Working Memory Capacity: A Prospective, Direct Observation Study. *BMJ Quality & Safety*, **27**, 655-663. <https://doi.org/10.1136/bmjqs-2017-007333>