

Evaluating Open-Source Generative AI in Biometry: A Retrospective Study on DeepSeek Performance for Intraocular Lens (IOL) Power Calculation

Giovanni Ottonelli, Giacomo De Rosa, Jacopo Celada Ballanti, Alessandro Gaeta,
Paolo Vinciguerra, Alessandra Di Maria

Submitted to: JMIR Formative Research
on: April 29, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 24

 Figures 25

 Figure 1..... 26

 Figure 2..... 27

 Figure 3..... 28

 Figure 4..... 29

Evaluating Open-Source Generative AI in Biometry: A Retrospective Study on DeepSeek Performance for Intraocular Lens (IOL) Power Calculation

Giovanni Ottonelli¹ MD; Giacomo De Rosa¹ MD; Jacopo Celada Ballanti¹ MD; Alessandro Gaeta²; Paolo Vinciguerra¹ MD; Alessandra Di Maria¹ MD

¹ IRCCS Humanitas Research Hospital Milan IT

² University of Genoa Genoa IT

Corresponding Author:

Giovanni Ottonelli MD

IRCCS Humanitas Research Hospital
Rozzano
Milan
IT

Abstract

Background: Intraocular lens (IOL) power calculation is crucial for achieving optimal refractive outcomes in cataract surgery. Several formulas have been developed to predict postoperative refraction, with varying degrees of accuracy. Recent advancements in artificial intelligence (AI) have led to the exploration of AI-driven tools as potential alternatives or adjuncts to traditional formulas. DeepSeek, a generative AI model, is one such tool that has been proposed for biometric calculations. However, its performance in predicting refractive outcomes compared to conventional methods remains unclear. This study aims to assess the accuracy of DeepSeek in predicting postoperative refraction and compare it with a widely used formula, the Barrett Universal II.

Objective: To evaluate the accuracy of DeepSeek, an open-source generative artificial intelligence (AI), in predicting postoperative refractive outcomes for intraocular lens (IOL) power calculation compared to the Barrett Universal II formula.

Methods: This study analyzed biometric data from 100 eyes of 50 patients who underwent cataract surgery between July 2024 and January 2025 at Humanitas Research Hospital in Milan, Italy. Only uncomplicated cases of emmetropia with Alcon AcrySof® SA60WF IOL implantation were included. Postoperative objective refraction was measured with Nidek TONOREF III. Predicted spherical equivalents were obtained using the Barrett Universal II formula and DeepSeek AI. Mean refractive error (MRE), standard deviation (SD), and cumulative frequency of target refractive error were calculated. A Wilcoxon signed-rank test was performed to assess statistical significance.

Results: The MRE was significantly lower for Barrett (0.262 D) than for DeepSeek (0.659 D) ($P < 0.0001$). The SD was 0.386 D for Barrett and 0.424 D for DeepSeek. Barrett achieved refractive accuracy within ± 0.25 D in 37% of cases, ± 0.50 D in 75%, ± 0.75 D in 85%, and ± 1.0 D in 90%. DeepSeek performed worse at each threshold, reaching 10%, 30%, 55%, and 80%, respectively.

Conclusions: DeepSeek AI demonstrated significantly higher refractive errors and lower predictive accuracy than the Barrett Universal II formula. Although AI is a valuable tool already integrated into some biometric formulas, current widely accessible generative models remain inadequate for accurate biometric calculations, emphasizing the need for human-validated formulas to ensure optimal refractive outcomes in cataract surgery.

(JMIR Preprints 29/04/2025:76717)

DOI: <https://doi.org/10.2196/preprints.76717>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://www.jmir.org/>



Original Manuscript

Title:

“Evaluating Open-Source Generative AI in Biometry: A Retrospective Study on DeepSeek Performance for Intraocular Lens (IOL) Power Calculation”

Abstract

Purpose: To evaluate the accuracy of DeepSeek, an open-source generative artificial intelligence (AI), in predicting postoperative refractive outcomes for intraocular lens (IOL) power calculation compared to the Barrett Universal II formula.

Setting: Tertiary Ophthalmology Center.

Design: Retrospective and comparative study.

Methods: This study analyzed biometric data from 100 eyes of 50 patients who underwent cataract surgery between July 2024 and January 2025 at Humanitas Research Hospital in Milan, Italy. Only uncomplicated cases of emmetropia with Alcon AcrySof® SA60WF IOL implantation were included. Postoperative objective refraction was measured with Nidek TONOREF III. Predicted spherical equivalents were obtained using the Barrett Universal II formula and DeepSeek AI. Mean refractive error (MRE), standard deviation (SD), and cumulative frequency of target refractive error were calculated. A Wilcoxon signed-rank test was performed to assess statistical significance.

Results: The MRE was significantly lower for Barrett (0.262 D) than for DeepSeek (0.659 D) ($P < 0.0001$). The SD was 0.386 D for Barrett and 0.424 D for DeepSeek. Barrett achieved refractive accuracy within ± 0.25 D in 37% of cases, ± 0.50 D in 75%, ± 0.75 D in 85%, and ± 1.0 D in 90%. DeepSeek performed worse at each threshold, reaching 10%, 30%, 55%, and 80%, respectively.

Conclusions: DeepSeek AI demonstrated significantly higher refractive errors and lower predictive accuracy than the Barrett Universal II formula. Although AI is a valuable tool already integrated into some biometric formulas, current widely accessible generative models remain inadequate for accurate biometric calculations, emphasizing the need for human-validated formulas to ensure

optimal refractive outcomes in cataract surgery.

Keywords

Generative artificial intelligence, artificial intelligence, DeepSeek, ChatGPT, ophthalmology, intraocular lens, IOL, spheric equivalent, biometry, cataract, cataract surgery

Introduction

Biometry in ophthalmology refers to the precise measurement of ocular anatomical features, notably corneal power and axial length, which are essential for determining the appropriate intraocular lens (IOL) power in cataract surgery ¹. By precisely capturing these variables, surgeons aim to minimize postoperative refractive errors and reduce reliance on spectacles or contact lenses ². Traditional ultrasound methods have largely been superseded by optical biometry techniques, such as partial coherence interferometry and low-coherence reflectometry, which allow for more accurate and reproducible AL measurements ³. Modern IOL power formulas, including Haigis ⁴, Holladay 2 ⁵, and Barrett Universal II ⁶, incorporate additional parameters, such as anterior chamber depth (ACD) or lens thickness (LT), to refine predictions of the effective lens position (ELP) and give accurate results. Nonetheless, certain clinical scenarios continue to pose challenges, particularly in eyes with unusual anatomy (e.g., extreme myopia or prior refractive surgery). These situations often result in unexpected outcomes when using classical IOL formulas ^{7,8}. To address these outliers, advanced ray-tracing models simulate light propagation through the cornea and IOL, offering more personalized calculations ⁹. Meanwhile, machine-learning approaches have started to augment or surpass conventional formulas by identifying subtle correlations in large clinical datasets ¹⁰.

On the other hand, generative artificial intelligence (AI) has emerged as a transformative tool in medicine, offering innovative solutions for diagnostics, treatment planning, and personalized care ¹¹. By leveraging advanced machine learning models, such as generative adversarial networks (GANs) and large language models (LLMs), generative AI can synthesize complex medical data, including

imaging, genomic information, and clinical notes, to assist healthcare professionals in decision-making^{12,13}.

AI has become a powerful tool in ophthalmology, particularly for diagnosing and managing eye diseases. AI algorithms, especially those based on deep learning, excel at detecting conditions like diabetic retinopathy, age-related macular degeneration (AMD), and glaucoma from fundus oculi images, often matching or surpassing human experts in accuracy^{14,15}. AI also supports telemedicine by enabling remote monitoring of chronic conditions like glaucoma, making eye care more accessible¹⁶. Beyond diagnostics, AI aids in predicting postoperative outcomes in intraocular collamer lens (ICL) implants, refractive phacoemulsification, and cataract surgery by analyzing preoperative biometric data, helping surgeons select the optimal intraocular lens (IOL) power for better visual outcomes^{17,18}. Additionally, AI approaches have been applied to complex eye biometry such as short axial length eyes¹⁹.

However, challenges remain, including the need for diverse datasets, data privacy concerns, and seamless integration into clinical workflows. Ethical and regulatory considerations are also critical to ensure safe and equitable use¹⁴.

In this rapidly evolving landscape, *DeepSeek* has recently drawn attention as an open-source deep learning-based generative artificial intelligence²⁰. Since its birth in 2023, and even more since the launch of DeepSeek's DeepThink (R1), this generative AI has been capable of advanced reasoning at lower training costs than many rivals^{21,22}. Multiple research groups are now investigating how to incorporate DeepSeek within secure data environments, aiming to accelerate discovery while mitigating biases in training data^{23,24}.

In the context of healthcare, evidence regarding DeepSeek's applications remains limited. This is primarily due to the very recent launch of its latest version (released on January 2025) and the competitive landscape dominated by earlier generative AI systems, such as ChatGPT, which have already established a significant presence in the field. While these earlier systems have been

extensively studied and applied across various medical domains, as research progresses, DeepSeek's unique capabilities, such as advanced reasoning and cost-efficiency, may offer distinct advantages for precision-driven healthcare applications [26]. Consequently, we aimed to explore the application of DeepSeek in the field of ocular biometry, where AI-driven tools hold the potential for refining lens constants, predicting postoperative IOL positioning, and optimizing IOL power calculations. By utilizing DeepSeek's advanced reasoning capabilities, we sought to evaluate its performance as a non-inferior tool for standard-eye IOL power calculations.

Methods

This retrospective study analyzed 100 biometric datasets from 100 eyes of 50 patients who underwent uncomplicated cataract surgery at Humanitas Research Hospital in Milan, Italy, between July 2024 and January 2025. To ensure comparability with calculations, only patients planned for and implanted with an intraocular lens (IOL) power targeting emmetropia were included. Emmetropia was defined by choosing the first negative IOL power closest to a spherical equivalent of 0.00 diopters (D). The exclusion criterion were a planned postoperative refractive target other than emmetropia, complications during surgery and previous refractive surgery.

Preoperative biometric data were obtained using the Heidelberg Engineering Anterior® biometry tool ²⁶. The following parameters were collected for each patient: age, gender (M/F), eye (right/left, OD/OS), keratometry values (K1 and K2, in diopters [D]), anterior chamber depth (ACD, in millimeters [mm]), axial length (AL, in mm), lens thickness (LT, in mm), and white-to-white distance (WTW, in mm). *Table 1* shows a summary of biometric data.

Table 1 – Biometric data demographics

The table shows the mean and standard deviation (SD, the minimum and the maximum value of the reported parameters. AXL = axial length; K1 = flat keratometry; K2 = steep keratometry; ACD =

anterior chamber depth; LT = lens thickness; WTW = white-to-white.

	Mean $\pm$ SD	Minimum	Maximum
AXL (mm)	23.59 $\pm$ 1.29	21.10	26.96
K1 (mm)	42.86 $\pm$ 1.77	36.20	44.93
K2 (mm)	44.14 $\pm$ 1.60	37.30	45.89
ACD (mm)	3.08 $\pm$ 0.51	2.30	4.06
LT (mm)	4.72 $\pm$ 0.53	3.23	5.28
WTW (mm)	12.02 $\pm$ 0.34	11.10	12.70
Implanted IOL (D)	21.89 $\pm$ 2.66	16.00	24.50
Postoperative Spherical Equivalent (D)	0.14 $\pm$ 0.41	-1.00	1.00

We chose the Barrett Universal II formula for its well-known overall precision compared to other formulas^{27,28}.

All the surgeries were performed using a temporal clear corneal main port of 2.2mm and a 1.0 mm paracentesis. The device used for phacoemulsification was the Alcon Centurion and all the implanted IOLs were Alcon AcrySof® SA60WF.

The final objective refraction after cataract surgery was measured with Nidek TONOREF III.

To extract IOL power calculations with DeepSeek AI the following command was entered into the “Message DeepSeek” tool for each eye:

“Hi! I’m an ophthalmic surgeon and I need you to assist me with intraocular lens (IOL) power calculation data analysis. Given the following patient:

- *Gender (Male/Female)*
- *Eye (OD/OS)*
- *AXL: XX.XX mm (e.g. 24.89 mm)*

- *K1: XX.XX D (e.g. 45.67 D)*
- *K2: XX.XX D (e.g. 43.12 D)*
- *ACD: X.XX mm (e.g. 4.34 mm)*
- *LT: X.XX mm (e.g. 3.63 mm)*
- *WTW: XX.XX mm*
- *Implanted IOL optic A constant (Alcon AcrySof® SN60WF): 119.0*

TASK: I Want you to tell me the final spheric equivalent I would get after implanting a +XX.XX D IOL in this specific eye. Just give me the final number.

NOTE: the result may also be hyperopic if you calculated so and does not need to be round to 0.05 values

To minimize DeepSeek calculation biases, the DeepSeek tool was instructed to provide only the numerical result without explanatory text, and both the DeepThink (R1) and Web Search modalities were activated.

Postoperative objective refractive outcomes were compared between the predicted values from the Barrett Universal II formula and those obtained using DeepSeek. This comparison aimed to determine which method more accurately predicted the final refractive target. The mean refractive error (MRE, defined as the arithmetic difference between the postoperative spherical equivalent and the predicted spherical equivalent) and the cumulative frequency of refractive target error were calculated. The cumulative frequency analysis assessed the proportion of eyes achieving specific levels of refractive error relative to the desired target refraction.

The primary endpoint of the study was to evaluate the accuracy of DeepSeek calculations in predicting the final refractive outcome compared to the Barrett Universal II formula by comparing the mean of the MRE of the two methods.

Figures and graphs were created using the Journal of Cataract and Refractive Surgery online Excel

tool. Statistical analysis was conducted using the GraphPad® online calculator, with significance set at a p-value of less than 0.05²⁹. Given the paired nature of the non-normally distributed data, a Wilcoxon signed-rank test was performed to assess whether the differences between the two methods were statistically significant. This test, commonly used to compare means of non-normally distributed data from related groups, was applied to the absolute prediction errors, defined as the absolute differences between the actual postoperative spherical equivalent and the predicted values from each method. A p-value of <0.05 indicated a statistically significant difference between the two prediction methods, confirming a discrepancy in their performance.

A heteroscedasticity analysis was also conducted using an F-test to compare variances between the Barrett Universal II and DeepSeek methods.

It is important to state that DeepSeek was used without specific fine-tuning on ophthalmology datasets. It operated as a generalist AI model, and no training optimization specific to biometry was performed. Additionally, IOL constants were not optimized to nullify MRE for DeepSeek predictions, as DeepSeek did not allow adjustment of lens constants within its structure.

Results

In **Figure 1** and **Figure 2**, the postoperative spherical equivalent (SE) is plotted against the predicted SE for the Barrett Universal II formula and the DeepSeek method, respectively. These visual representations illustrate the distribution of refractive outcomes relative to the predictions made by each approach. A comparative analysis of key descriptive statistics reveals that the Barrett Universal II formula demonstrated a mean refractive error (MRE) of 0.262 diopters (D), whereas DeepSeek exhibited a significantly higher MRE of 0.659 D. The standard deviation (SD) associated with Barrett was 0.386 D, indicating the degree of variability in its predictions, while DeepSeek presented a comparable SD of 0.424 D. The maximum refractive error observed with the Barrett Universal II formula was 1.23 D, while DeepSeek exhibited a larger maximum error of 1.695 D. The minimum

recorded refractive error for Barrett was -0.015 D, whereas for DeepSeek, it was -0.02 D.

A visual comparison of the distribution of refractive prediction errors for both methods is provided in **Figure 3**, which employs a box plot format. This representation highlights the differences in the median values, interquartile ranges, and presence of outliers between the two methods. To rigorously assess the absolute prediction errors, a Wilcoxon signed-rank test was performed. The analysis yielded a highly statistically significant p-value of less than 0.0001 ($P < 0.0001$), confirming a significant difference in the predictive performance of the two methods.

Figure 4 illustrates the cumulative frequency of refractive errors at various thresholds, highlighting the superior predictive accuracy of the Barrett Universal II formula. Within ± 0.25 D, Barrett accurately predicted outcomes in 37% of eyes, compared to 10% for DeepSeek. This trend persisted across broader thresholds, with Barrett achieving 75% accuracy within ± 0.50 D versus 30% for DeepSeek, 85% within ± 0.75 D versus 55%, and 90% within ± 1.0 D compared to DeepSeek's 80%. In conclusion, the findings of this study indicate that DeepSeek did not achieve non-inferiority when compared with the Barrett Universal II formula for intraocular lens (IOL) power calculations. The statistical analyses and cumulative frequency data consistently demonstrate that the Barrett Universal II formula provided more accurate and predictable refractive outcomes in this cohort.

Variability between the two prediction methods was compared using an F-test for equality of variances. The analysis produced an F value of 1.208 ($P = 0.45$) indicating that the spread of prediction errors was statistically comparable between methods.

Discussion

Our findings indicate that DeepSeek, as an open-source generative artificial intelligence, demonstrated significantly lower accuracy in intraocular lens (IOL) power prediction compared to the Barrett Universal II formula. The mean refractive error (MRE) for DeepSeek was markedly

higher (0.659 D vs. 0.262 D), with a statistically significant difference ($P < .001$). These results confirm that DeepSeek, in its current form, is inadequate for reliable biometric calculations in cataract surgery. Previous studies have highlighted the potential of AI in ophthalmology, particularly in diagnostic applications and decision-support tools for complex biometric cases^{6,14}. However, large language models such as DeepSeek require extensive domain-specific fine-tuning to meet the rigorous accuracy standards necessary for clinical application¹¹. The performance gap observed in this study suggests that DeepSeek lacks the precision required to replace established biometric formulas, particularly in standard IOL calculations. Recent studies have further emphasized the importance of validating AI models through rigorous clinical testing and fine-tuning, particularly when incorporating AI into systems like IOL calculations in cataract surgery³⁰.

Furthermore, our study underscores the importance of validated, human-supervised methodologies in AI integration for medical applications. The cumulative frequency analysis revealed that Barrett Universal II significantly outperformed DeepSeek at all refractive error thresholds. Within ± 0.25 D, Barrett achieved 37% accuracy, compared to 10% for DeepSeek, and this disparity persisted at ± 0.50 D (75% vs. 30%) and ± 1.0 D (90% vs. 80%). This trend is consistent with prior research emphasizing the robustness of traditional biometric formulas over purely AI-driven approaches⁷. Several studies have reinforced the need for a human-centered, collaborative approach when applying AI in ophthalmology, ensuring AI models remain adjunctive tools rather than independent decision-makers. AI should function as a support tool for human expertise, such as in the management of complex cases or rare diseases³¹. A potential explanation for DeepSeek's underperformance lies in its inherent training limitations. Unlike established formulas optimized for ocular biometric calculations, DeepSeek is a generalist AI model not specifically trained on ophthalmic datasets. Additionally, generative AI may introduce biases due to dataset selection and input variability, which could further compromise its accuracy in a clinical setting²⁵. The application of AI in ophthalmology, especially for disease prediction and management, is highly reliant on

training data quality. Current AI models, if not properly trained with domain-specific data, may fail to meet the precision required for clinical practice. This highlights the importance of training AI with large, accurate datasets that are specifically curated for ophthalmic diseases³².

Interestingly, DeepSeek's development highlights a broader shift in the AI landscape. The model's efficiency in reasoning and problem-solving, despite using fewer computational resources than its competitors, signals a potential paradigm shift away from the traditional scaling law of AI, which prioritizes larger datasets and increased computational power²². The rapid development of DeepSeek R1, achieved with a small research team, has demonstrated that innovative algorithmic strategies can match or even surpass larger, more resource-intensive models¹². Such innovations in AI algorithms point to the growing potential of AI to revolutionize medical practices, including ophthalmology, through more efficient, scalable tools that are less dependent on resource-intensive infrastructures³³. Furthermore, DeepSeek's implementation of Mixture of Experts, reinforcement learning, and FP8 mixed precision frameworks has allowed it to achieve notable efficiency gains while maintaining competitive performance across benchmarks⁹. In ophthalmology, AI models like DeepSeek could potentially aid in predicting systemic parameters, helping doctors not only assess eye health but also detect early signs of systemic diseases that may manifest through ophthalmic imaging³⁴.

Another distinguishing feature of DeepSeek is its reasoning process. Unlike traditional LLMs, which often function as black boxes, DeepSeek leverages chain-of-thought (CoT) reasoning to break down complex tasks into logical steps. This characteristic could theoretically benefit medical applications, such as differential diagnoses, where multi-step reasoning is essential¹⁸. However, in our study, this reasoning approach did not translate into improved accuracy for biometric predictions, suggesting that domain-specific adaptations are still required for effective integration in ophthalmology.

Despite its limitations, generative AI still holds promise as an adjunct tool rather than a standalone solution. Future research should focus on integrating AI with traditional biometric formulas to harness the pattern recognition strengths of deep learning while maintaining the reliability of

validated formulas. Moreover, prospective studies should explore AI fine-tuning techniques tailored to ophthalmology, ensuring data governance and compliance with clinical safety standards ²¹. As AI continues to evolve in ophthalmology, the focus must shift toward refining its integration with existing clinical practices and ensuring that AI-driven tools remain compatible with medical ethics and clinical governance ³⁵.

While our study provides valuable insights into the performance of DeepSeek in IOL power prediction, several limitations should be acknowledged. First, we exclusively analyzed cases targeting emmetropia, which may limit the generalizability of our findings to other refractive targets, such as myopia. Additionally, preoperative refraction was not included as a parameter, potentially reducing the ability to fully understand individual variability in postoperative outcomes. Our study also lacked a detailed longitudinal assessment of postoperative refractive trends, which could have provided further insight into the long-term predictive accuracy of the AI model.

Another significant limitation is the exclusion of eyes with extreme biometric values, such as those with high myopia or hyperopia, which are particularly challenging for IOL power calculations. Evaluating DeepSeek's performance in these cases is crucial to assess its broader applicability. Furthermore, we tested only one version of DeepSeek using a single interface, which may not account for improvements in newer iterations or alternative AI models. Lastly, technical challenges, including frequent server congestion, resulted in slow data retrieval, with some calculations taking up to a week. This raises concerns regarding the practicality of real-time clinical implementation. In addition, the lack of IOL constant optimization for DeepSeek predictions, the small sample size, and the absence of a validation cohort further limit the generalizability of the findings. Future studies should address these challenges by refining generative AI methodologies and incorporating a wider range of biometric profiles.

In conclusion, while AI is rapidly evolving and has demonstrated utility in various domains, our study highlights the necessity for rigorous validation before generative models can be deployed in

critical clinical applications. DeepSeek current performance does not meet the threshold for reliable biometric predictions, reinforcing the continued importance of established IOL power calculation methods.

Blinded Acknowledgments

The authors acknowledge those who contributed to this research.

Value statement

What was known

- AI has been increasingly integrated into biometric formulas to enhance intraocular lens (IOL) power calculations.
- Traditional formulas, such as Barrett Universal II, have demonstrated high accuracy in predicting postoperative refractive outcomes.
- The accuracy and reliability of generative AI models, such as DeepSeek, for biometric calculations remain unproven.

What this paper adds

- DeepSeek demonstrated significantly lower accuracy in IOL power prediction compared to the Barrett Universal II formula.
- Generative AI models currently fail to achieve non-inferiority in biometric calculations, with higher mean refractive errors and greater variability.
- Despite AI's advancements, human-validated formulas remain essential for achieving optimal refractive outcomes in cataract surgery.

Blinded conflict of Interest statement

The authors declare no conflicts of interest related to this study.

References

1. Olsen T. Calculation of intraocular lens power: a review. *Acta Ophthalmol Scand.* 2007;85(5):472-485. doi:10.1111/J.1600-0420.2007.00879.X
2. Norrby S. Sources of error in intraocular lens power calculation. *J Cataract Refract Surg.* 2008;34(3):368-376. doi:10.1016/J.JCRS.2007.10.031
3. Montés-Micó R, Carones F, Buttacchio A, Ferrer-Blasco T, Madrid-Costa D. Comparison of immersion ultrasound, partial coherence interferometry, and low coherence reflectometry for ocular biometry in cataract patients. *J Refract Surg.* 2011;27(9):665-671. doi:10.3928/1081597X-20110202-01
4. Haigis W, Hoffer KJ, Hoffer KJ. The Haigis Formula. *Essentials in Ophthalmology.* Published online 2024:625-636. doi:10.1007/978-3-031-50666-6_41
5. Ioannides A. Holladay 2 formula. *Ophthalmology.* 2011;118(12):2525. doi:10.1016/j.ophtla.2011.07.023
6. Barrett GD. Barrett Formulas: Strategies to Improve IOL Power Prediction. *Essentials in Ophthalmology.* Published online 2024:577-592. doi:10.1007/978-3-031-50666-6_37
7. Wang L, Shirayama M, Ma XJ, Kohnen T, Koch DD. Optimizing intraocular lens power calculations in eyes with axial lengths above 25.0 mm. *J Cataract Refract Surg.* 2011;37(11):2018-2027. doi:10.1016/J.JCRS.2011.05.042
8. Haigis W. Intraocular lens calculation after refractive surgery for myopia: Haigis-L formula. *J Cataract Refract Surg.* 2008;34(10):1658-1663. doi:10.1016/J.JCRS.2008.06.029
9. Preussner PR, Wahl J, Lahdo H, Dick B, Findl O. Ray tracing for intraocular lens calculation. *J Cataract Refract Surg.* 2002;28(8):1412-1419. doi:10.1016/S0886-3350(01)01346-3
10. Mori Y, Yamauchi T, Tokuda S, Minami K, Tabuchi H, Miyata K. Machine learning adaptation of intraocular lens power calculation for a patient group. *Eye and Vision.* 2021;8(1):42.

doi:10.1186/S40662-021-00265-Z

11. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol.* 2017;2(4):230-243. doi:10.1136/SVN-2017-000101
12. DeepSeek-AI, :, Bi X, et al. DeepSeek LLM: Scaling Open-Source Language Models with Longtermism. Published online January 5, 2024. Accessed March 11, 2025. <https://arxiv.org/abs/2401.02954v1>
13. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *Commun ACM.* 2020;63(11):139-144. doi:10.1145/3422622
14. Li Z, Wang L, Wu X, et al. Artificial intelligence in ophthalmology: The path to the real-world clinic. *Cell Rep Med.* 2023;4(7):101095. doi:10.1016/J.XCRM.2023.101095
15. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA.* 2016;316(22):2402-2410. doi:10.1001/JAMA.2016.17216
16. Li JPO, Liu H, Ting DSJ, et al. Digital technology, tele-medicine and artificial intelligence in ophthalmology: A global perspective. *Prog Retin Eye Res.* 2021;82. doi:10.1016/J.PRETEYERES.2020.100900
17. Rampat R, Debellemanni re G, Gatinel D, Ting DSJ. Artificial intelligence applications in cataract and refractive surgeries. *Curr Opin Ophthalmol.* 2024;35(6):480-486. doi:10.1097/ICU.0000000000001090
18. Russo A, Filini O, Savini G, et al. Predictability of the vault after implantable collamer lens implantation using OCT and artificial intelligence in White patient eyes. *J Cataract Refract Surg.* 2023;49(7):724-731. doi:10.1097/J.JCRS.0000000000001182
19. Savini G, Taroni L, Hoffer KJ. Comment on: Efficacy of segmented axial length and artificial intelligence approaches to intraocular lens power calculation in short eyes. *J Cataract Refract Surg.* 2024;50(2):195. doi:10.1097/J.JCRS.0000000000001358

20. DeepSeek Artificial Intelligence. 2023. Accessed March 11, 2025. <https://deepseek.ai/>
21. Mondillo G, Colosimo S, Perrotta A, Frattolillo V, Masino M. Comparative Evaluation of Advanced AI Reasoning Models in Pediatric Clinical Decision Support: ChatGPT O1 vs. DeepSeek-R1. *medRxiv*. Published online January 28, 2025;2025.01.27.25321169. doi:10.1101/2025.01.27.25321169
22. Kayaalp ME, Prill R, Sezgin EA, Cong T, Królikowska A, Hirschmann MT. DeepSeek versus ChatGPT: Multimodal artificial intelligence revolutionizing scientific discovery. From language editing to autonomous content generation-Redefining innovation in research and practice. *Knee Surg Sports Traumatol Arthrosc*. Published online 2025. doi:10.1002/KSA.12628
23. Qin Z, Gui Q, Bian M, et al. Multimodal Human-AI Synergy for Medical Imaging Quality Control: A Hybrid Intelligence Framework with Adaptive Dataset Curation and Closed-Loop Evaluation. Published online March 10, 2025. Accessed March 11, 2025. <https://arxiv.org/abs/2503.07032v1>
24. Qiu P, Wu C, Liu S, et al. Quantifying the Reasoning Abilities of LLMs on Real-world Clinical Cases. Published online March 6, 2025. Accessed March 11, 2025. <https://arxiv.org/abs/2503.04691v2>
25. Temsah A, Alhasan K, Altamimi I, et al. DeepSeek in Healthcare: Revealing Opportunities and Steering Challenges of a New Open-Source Artificial Intelligence Frontier. *Cureus*. 2025;17(2):e79221. doi:10.7759/CUREUS.79221
26. Heidelberg Engineering. ANTERION – multimodal imaging platform for the anterior segment. Accessed March 11, 2025. <https://business-lounge.heidelbergengineering.com/us/en/products/anterion/>
27. Cooke DL, Cooke TL. Comparison of 9 intraocular lens power calculation formulas. *J Cataract Refract Surg*. 2016;42(8):1157-1164. doi:10.1016/J.JCRS.2016.06.029
28. Kane JX, Van Heerden A, Atik A, Petsoglou C. Intraocular lens power formula accuracy: Comparison of 7 formulas. *J Cataract Refract Surg*. 2016;42(10):1490-1500.

doi:10.1016/J.JCRS.2016.07.021

29. GraphPad. Accessed March 22, 2025. <https://www.graphpad.com/quickcalcs/ttest1/>
30. Cappellani F, Card KR, Shields CL, Pulido JS, Haller JA. Reliability and accuracy of artificial intelligence ChatGPT in providing information on ophthalmic diseases and management to patients. *Eye* 2024 38:7. 2024;38(7):1368-1373. doi:10.1038/s41433-023-02906-0
31. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *British Journal of Ophthalmology*. 2019;103(2):167-175. doi:10.1136/BJOPHTHALMOL-2018-313173
32. Betzler BK, Rim TH, Sabanayagam C, Cheng CY. Artificial Intelligence in Predicting Systemic Parameters and Diseases From Ophthalmic Imaging. *Front Digit Health*. 2022;4:889445. doi:10.3389/FDGTH.2022.889445/PDF
33. Lin WC, Chen JS, Chiang MF, Hribar MR. Applications of Artificial Intelligence to Electronic Health Record Data in Ophthalmology. *Transl Vis Sci Technol*. 2020;9(2):13-13. doi:10.1167/TVST.9.2.13
34. Tan YY, Kang HG, Lee CJ, et al. Prognostic potentials of AI in ophthalmology: systemic disease forecasting via retinal imaging. *Eye and Vision*. 2024;11(1):1-18. doi:10.1186/S40662-024-00384-3/TABLES/3
35. Akkara J, Kuriakose A. Role of artificial intelligence and machine learning in ophthalmology. *Kerala Journal of Ophthalmology*. 2019;31(2):150. doi:10.4103/KJO.KJO_54_19

Figures legends

Figure 1 – Spherical Equivalent (SEQ) Attempted vs Achieved with Barrett II Universal formula

The scatter plot depicts achieved spherical equivalent refraction (D) against attempted refraction (D), with data points representing individual eyes. Accuracy is quantified by linear boundaries: $\pm 0.5D$

(green lines) and $\pm 1.0D$ (pink lines), indicating the distribution of refractive errors relative to the predicted target. Regions labeled “Overcorrected” and “Undercorrected” denote deviations from intended refractive outcomes.

Figure 2 – Spherical Equivalent (SEQ) Attempted vs Achieved with DeepSeek AI

The scatter plot depicts achieved spherical equivalent refraction (D) against attempted refraction (D), with data points representing individual eyes. Accuracy is quantified by linear boundaries: $\pm 0.5D$ (green lines) and $\pm 1.0D$ (pink lines), indicating the distribution of refractive errors relative to the predicted target. Regions labeled “Overcorrected” and “Undercorrected” denote deviations from intended refractive outcomes.

Figure 3 – Box Plot

The figure compares refractive prediction errors (SEQ) between Barrett II Universal and DeepSeek AI using box plots. DeepSeek exhibits a higher median error of approximately 0.6 D (mean 0.7 D) compared to Barrett II's 0.2 D median and mean, suggesting a systematic overestimation. Barrett II demonstrates greater variability with an interquartile range (IQR) of 0.6 D, spanning 0.1 D to 0.7 D, and a wider range of -0.5 D to 1.0 D (excluding outliers). DeepSeek IQR is 0.4 D (0.5 D to 0.9 D) with a range of -0.3 D to 1.5 D.

Figure 4 – Cumulative Frequency

The chart compares cumulative percentages of eyes achieving specific refractive errors using Barrett II Universal and DeepSeek AI. At 0.25D, Barrett II shows approximately 38% achieving this accuracy, compared to DeepSeek's 10%. At 0.50D, Barrett II reaches roughly 75%, while DeepSeek is around 30%. At 1.00D, Barrett II attains nearly 95% accuracy, DeepSeek around 80%. Both formulas approach 100% accuracy at 1.75D. Barrett II consistently outperforms DeepSeek across all

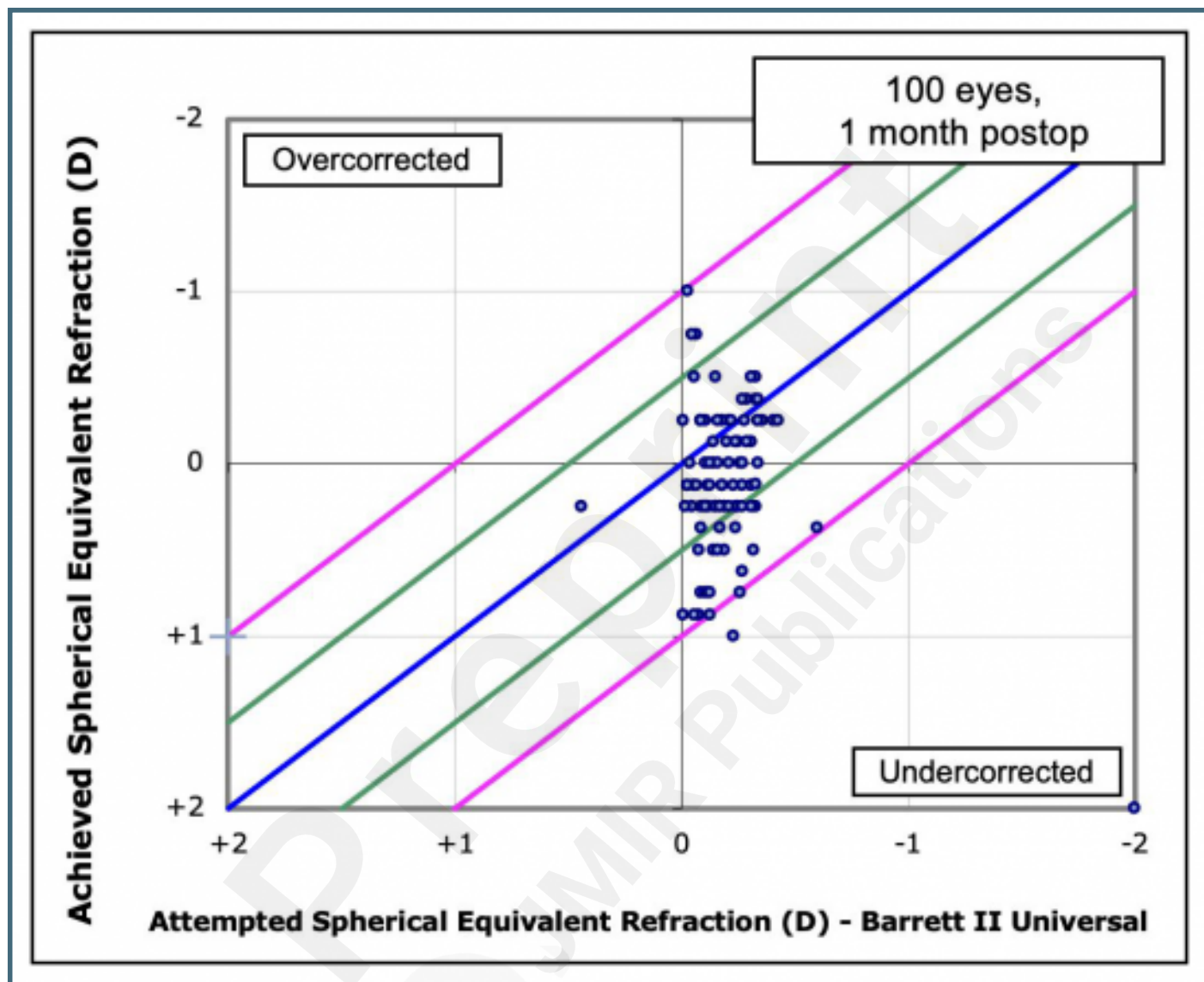
reported error ranges.



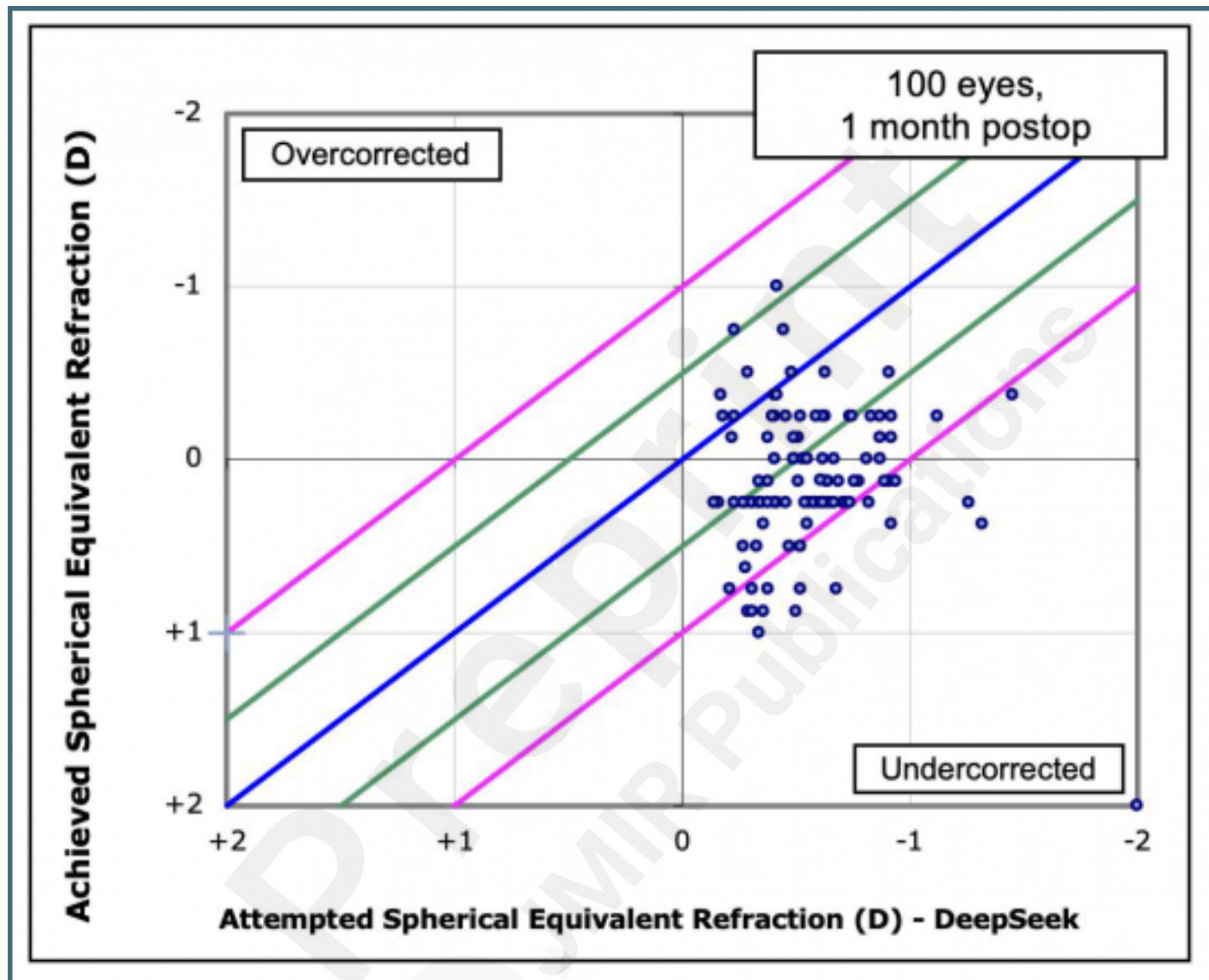
Supplementary Files

Figures

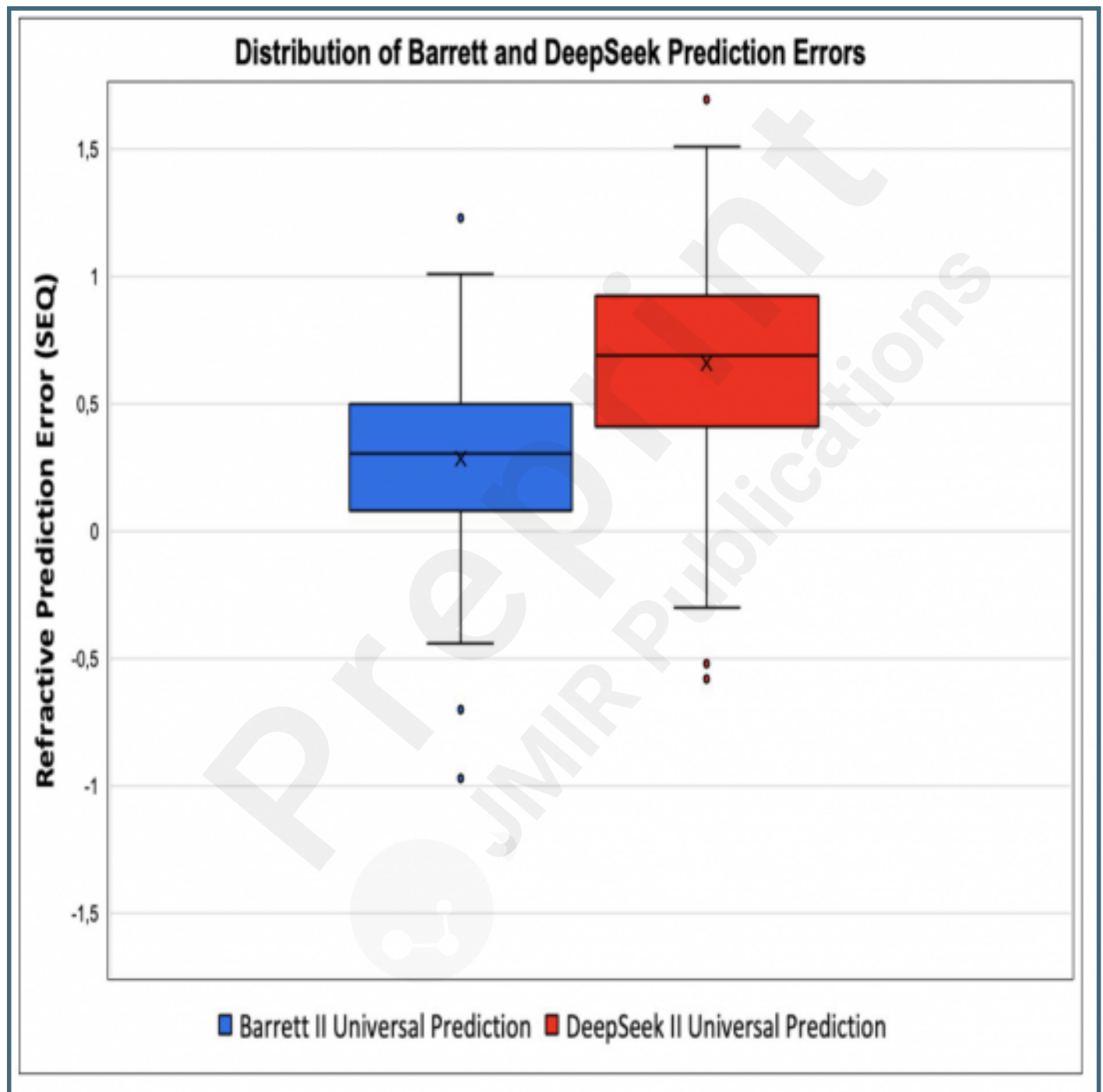
The scatter plot depicts achieved spherical equivalent refraction (D) against attempted refraction (D), with data points representing individual eyes. Accuracy is quantified by linear boundaries: $\pm 0.5\text{D}$ (green lines) and $\pm 1.0\text{D}$ (pink lines), indicating the distribution of refractive errors relative to the predicted target. Regions labeled “Overcorrected” and “Undercorrected” denote deviations from intended refractive outcomes.



The scatter plot depicts achieved spherical equivalent refraction (D) against attempted refraction (D), with data points representing individual eyes. Accuracy is quantified by linear boundaries: $\pm 0.5\text{D}$ (green lines) and $\pm 1.0\text{D}$ (pink lines), indicating the distribution of refractive errors relative to the predicted target. Regions labeled “Overcorrected” and “Undercorrected” denote deviations from intended refractive outcomes.



The figure compares refractive prediction errors (SEQ) between Barrett II Universal and DeepSeek AI using box plots. DeepSeek exhibits a higher median error of approximately 0.6 D (mean 0.7 D) compared to Barrett II's 0.2 D median and mean, suggesting a systematic overestimation. Barrett II demonstrates greater variability with an interquartile range (IQR) of 0.6 D, spanning 0.1 D to 0.7 D, and a wider range of -0.5 D to 1.0 D (excluding outliers). DeepSeek IQR is 0.4 D (0.5 D to 0.9 D) with a range of -0.3 D to 1.5 D.



The chart compares cumulative percentages of eyes achieving specific refractive errors using Barrett II Universal and DeepSeek AI. At 0.25D, Barrett II shows approximately 38% achieving this accuracy, compared to DeepSeek's 10%. At 0.50D, Barrett II reaches roughly 75%, while DeepSeek is around 30%. At 1.00D, Barrett II attains nearly 95% accuracy, DeepSeek around 80%. Both formulas approach 100% accuracy at 1.75D. Barrett II consistently outperforms DeepSeek across all reported error ranges.

