Evaluation of ChatGPT's Responses to Google Searches About Dry Eye

Halil İbrahim Sönmezoğlu Sr, Büşra Güner Sönmezoğlu

Submitted to: JMIR AI
on: May 08, 2024

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 it's 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 expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Evaluation of ChatGPT's Responses to Google Searches About Dry Eye

Halil ?brahim Sönmezo?lu Sr[1] MD; Bü?ra Güner Sönmezo?lu[2] MD

[1]Hendek State Hospital TURKISH MINISTRY OF HEALTH SAKARYA TR
[2]Serdivan State Hospital TURKISH MINISTRY OF HEALTH SAKARYA TR

Corresponding Author:
Halil ?brahim Sönmezo?lu Sr MD
Hendek State Hospital
TURKISH MINISTRY OF HEALTH
Hendek
SAKARYA
TR

Abstract

Background: ChatGPT may have the potential to provide detailed information in the field of health and even dry eye

Objective: This study was to evaluate the text's quality, readability, and comprehensibility of the content generated by the ChatGPT about the most frequently searched queries on Google about dry eye disease (DED).

Methods: The research employed Google Trends to discover the most commonly searched terms associated with DED. These identified keywords were then entered into ChatGPT, and the generated responses were evaluated for quality using the Ensuring Quality Information for Patients tool (EQIP). The readability of the content was measured using both Flesch-Kincaid Grade Level (FKGL) and Flesch-Kincaid Reading Ease (FKRE) parameters

Results: The most commonly searched phrases were "eye drops," "dry eyes," and "dry eye drops." The countries that showed the greatest interest in these topics were the United States of America, Ireland, and the United Kingdom. The statistical analysis uncovered substantial concerns regarding the readability and comprehension of ChatGPT's written content about DED, indicating a necessity for enhancement. The low average EQIP value indicated the need to improve the quality and reliability of the content generated by ChatGPT.

Conclusions: The results of this indicated that the readability of ChatGPT's content on DED surpassed predefined standards but also highlighted concerns about its quality. Enhancing quality could be achieved by retraining the virtual intelligence with credible sources and verifying information through expert review can enhance the quality of the content.

(JMIR Preprints 08/05/2024:60357)
DOI: https://doi.org/10.2196/preprints.60357

Original Manuscript

Evaluation of ChatGPT's Responses to Google Searches About Dry Eye

Halil Ibrahim Sonmezoğlu[1], Busra Guner Sonmezoglu[2]

[1] Hendek State Hospital, Department of Ophthalmology, Sakarya, TURKEY

[2] Serdivan State Hospital, Department of Ophthalmology, Sakarya, TURKEY

ABSTRACT

Background: ChatGPT may have the potential to provide detailed information in the field of health and even dry eye

Objective: This study was to evaluate the text's quality, readability, and comprehensibility of the content generated by the ChatGPT about the most frequently searched queries on Google about dry eye disease (DED).

Methods: The research employed Google Trends to discover the most commonly searched terms associated with DED. These identified keywords were then entered into ChatGPT, and the generated responses were evaluated for quality using the Ensuring Quality Information for Patients tool (EQIP). The readability of the content was measured using both Flesch-Kincaid Grade Level (FKGL) and Flesch-Kincaid Reading Ease (FKRE) parameters.

Results: The most commonly searched phrases were "eye drops," "dry eyes," and "dry eye drops." The countries that showed the greatest interest in these topics were the United States of America, Ireland, and the United Kingdom. The statistical analysis uncovered substantial concerns regarding the readability and comprehension of ChatGPT's written content about DED, indicating a necessity for enhancement. The low average EQIP value indicated the need to improve the quality and reliability of the content generated by ChatGPT.

Conclusions: The results of this indicated that the readability of ChatGPT's content on

DED surpassed predefined standards but also highlighted concerns about its quality. Enhancing quality could be achieved by retraining the virtual intelligence with credible sources and verifying information through expert review can enhance the quality of the content.

Keywords: ChatGPT; Dry eye disease; Google Trends; Quality; Comprehension

INTRODUCTION

Dry eye disease is a common ocular disorder characterized by insufficient tear production or dysfunction, leading to an unstable tear film and damage to the surface of the eye. This complex condition causes irritation, vision problems, and instability in the tear film, potentially resulting in harm to the eye's surface due to elevated osmolarity and inflammation. It affects millions of people and can interfere with daily living and routine activities[1].

Developed by OpenAI, ChatGPT is an advanced online language model utilizing deep learning techniques to mimic human language patterns. As a part of the generative pre-training transformer family, it stands as one of the most prominent publicly available language models. Drawing from a vast text dataset, ChatGPT comprehends human language intricacies and nuances, enabling it to generate contextually relevant responses for diverse queries. Research indicates that ChatGPT has shown impressive capability in generating understandable answers to challenging clinical queries, positioning it as a valuable virtual aide for both patients and healthcare professionals[2,3].

While ChatGPT holds promise for various uses in the medical and healthcare sectors, such as identifying research topics, supporting clinical and laboratory diagnosis efforts, helping patients manage their health, it also presents challenges. These include a lack of personal and emotional interactions crucial for developing effective communication skills in healthcare education. There are concerns about data privacy as well as the risk of generating biased or inaccurate content. Consequently, there is a critical need to assess the quality, readability and comprehensibility of the output generated by ChatGPT - especially since these aspects are important for helping patients make optimal decisions and ensuring informed consents meet recommended standards[4,5].

While analyses of the quality, readability, and comprehensibility of ChatGPT-generated content on various medical conditions exist in the literature[6,7], to best of our knowledge, there is currently no analysis available for content related to DED. Therefore, to bridge this gap and to provide insights for future research, the aim of this study is to analyse quality, readability, and comprehensibility of ChatGPT-generated content about DED.

METHODS

The research took place at the Ophthalmology Clinic of Hendek State Hospital between January 6, 2024, and February 9, 2024. As the study did not involve any operations conducted within a living organism, there was no need for permission from the ethical committee[8].

Before conducting the searches, all personal web browser data was cleared as a precautionary measure. The tool Google Trends (https://trends.google.com/) was employed to identify the most commonly searched terms associated with DED. The search keywords were gathered from worldwide searches spanning from 2004 to the present day. Twenty-five highly relevant and frequently sought keywords about DED were recorded for subsequent analysis.

The identified keywords were sequentially input into the January 10 version of ChatGPT (https://chat.openai.com/), following the order in which they were originally searched. Prior to commencing these searches, all browser-related data was cleared. Additionally, a new account was specifically created for interacting with ChatGPT, ensuring clear differentiation. Each query was conducted on its own chat page to maintain separation and facilitate analysis. The collected responses were recorded for further analysis for quality, readability and comprehensibility.

The evaluation of the quality of the acquired texts was conducted using the Ensuring Quality Information for Patients (EQIP) tool. This comprehensive assessment covers various aspects of content, including coherence and writing excellence. It consists of 20 questions with response options 'yes,' 'no,' or 'does not apply.' Scores are calculated by tallying affirmative responses, which are then multiplied by a factor of 1. Responses marked as "partly" are also considered and multiplied by a factor of 0.5, while negative answers receive a multiplication factor of zero. The resulting values from these calculations are added together and adjusted based on the number of "not applicable" answers subtracted from 20 in the denominator to account for unanswered questions. The final value is converted to an EQIP score expressed as a percentage using a multiplication factor at 100%. Content that receives scores between 76% and 100% is classified as "well written," indicating high quality. Those scoring between 51% and 75% are categorized as displaying good quality with minor issues. Scores falling between 26% and 50% suggest serious quality issues, while content earning an EQIP score of 0%-25% is deemed to have severe quality issues[9,10].

The readability of the acquired texts was assessed using the FKGL and Flesch FKRE parameters. The FKGL calculation involves dividing the total number of words by the total number of sentences, multiplying by 0.39, then adding to it the result of dividing syllables by words, multiplied by 11.8, and subtracting 15.59 from this sum. This formula predicts the grade level needed for understanding based on sentence length and syllable count. A lower score indicates easier understanding while a higher one implies greater linguistic complexity. On the other hand, FKRE assesses document readability through a calculation involving average sentence length multiplied by 1.015 plus average number of syllables per word multiplied by 84.6; this difference is then subtracted from 206.835 to provide a reading ease score with higher numbers indicating easier readability and scores below 30 suggesting college graduate reading levels[11].

The statistical analysis was performed using SPSS version 27 (IBM, New York, USA). The normality of the data was evaluated using the Shapiro-Wilk test. Continuous data is typically represented using the mean value plus or minus the standard deviation, while categorical data is represented using frequency. Group differences were analyzed using the Kruskal-Wallis test to compute means. The significance threshold was 0.05, indicating a 95% confidence interval.

RESULTS

The most common keywords searched for in relation to DED were ‘Dry eyes,’Eye drops,’ and ‘Dry eye drops.’ Table 1 presents the top 25 keywords searched for concerning DED between 2004 and 2024, as identified by data from Google Trends.

As per the search data analysis, it has been observed that the United Kingdom, United States of America, and Ireland are the top three countries with the highest interest in DED, as demonstrated in Figure 1. Figure 2 highlights the increasing prevelence of DED since 2004. The mean FKRE score was 53.61±5.05, while the FKGL was 9.04±1.74(Table 2). A Kruskal-Wallis test revealed no statistically significant difference between the observed scores of EQIP, FKRE, and FKGL across different categories (p=0.644, p=0.994, p=0.828, respectively).

Table 1:Top 25 keywords searched about 'Dry eye' across countries: 2004-2024 (Based on Google Trends Data).

Rank	Keyword	Category of The Topic Based on EQIP
1	Dry eyes	Medication or product
2	Eye drops	Medication or product
3	Dry eye drops	Medication or product
4	Dry eye eye drops	Medication or product
5	Dry eyes eye drops	Medication or product
6	Dry eyes eye drops	Medication or product
7	Eye drops for dry eye	Medication or product

8	Eye drops for dry eyes	Medication or product
9	Drops for dry eyes	Medication or product
10	Dry eye symptoms	Condition or illness
11	Dry skin	Condition or illness
12	Dry eye syndrome	Condition or illness
13	Dry eye treatment	Medication or product
14	Best eye drops	Miscellaneous
15	Best dry eye drops	Medication or product
16	What is dry eye?	Condition or illness
17	Dry under eye	Miscellaneous
18	Eye cream	Medication or product
19	Best eye drops for dry eyes	Medication or product
20	Red eye	Miscellaneous
21	Best drops for dry dry eyes	Medication or product
22	Best eye drops for dry eyes	Medication or product
23	Dry eye causes	Miscellaneous
24	One dry eye	Miscellaneous
25	Eye pain	Miscellaneous

Tablo 2: Minimum-maximum,mean and standard deviation values of EQIP, The Flesch-Kincaid Reading Ease, The Flesch-Kincaid Grade Level, and percentages of categories.

Parameter	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**	**N**
EQIP	35	62,50	50,01	6,38315	25
The Flesch-Kincaid Reading Ease	44,70	62,70	53,61	5,05225	25
The Flesch-Kincaid Grade Level	6,80	10,40	9,0440	1,74244	25
		Percantage			
Categories	Medication or product	60%			15
	Miscellaneous	24%			6
	Condition or illness	16%			4

DISCUSSION

Based on this study,it was observed that the content produced by ChatGPT about DED was found to be above the recommended educational level. The results of the study also indicated that assessment of the ChatGPT-generated texts using EQIP fell within the category of "serious problems with quality." To best of our knowledge, this is the first study to examine both readability and quality of ChatGPT's responses regarding DED.

The study results indicate that the United Kingdom(UK), the United States, and Ireland had significant search interest in DED. This aligns with previous findings: Vidal-Rohr et al. reported a high prevalence of DED in the UK[12]; McCann et al. found an 8.1% occurrence in the US through meta-analysis[13]; Dana et al. noted a 5.28% prevalence of dry eye in the US with a recent upward trend[14]. There is evidence in the literature suggesting that children in the United States are utilizing digital devices more frequently at home to enhance their learning across various subjects such as reading, writing, mathematics, and science [15]. Moreover, the adoption of smart mobile devices has surged in the United Kingdom since

2010, surpassing personal computers in numbers[16]. Additionally, Ireland stands out as a front-runner among European Union states for spreading and employing digital Technologies[17]. It's worth noting that digital eye strain is recognized as a potential consequence of extensive screen use; this might contribute to the heightened prevalence of DED-related searches within these countries [18].

The top three most commonly queried terms were 'Dry eyes', 'Eye drops' and 'Dry eye drops.' The high occurrence of these search phrases suggests a significant desire to learn about DED symptoms and treatments. This highlights the importance of easily accessible and accurate information on these crucial topics. Healthcare professionals, professional health organizations, social media content creators, and virtual intelligence developers have the opportunity to address this need by creating high-quality and easily understandable content tailored to the specific areas of interest for patients.

The literature indicates that a higher readability level is associated with a reduced likelihood of the information being understandable to a large population segment [19]. Obtaining trustworthy and easily understandable information is crucial for patients seeking answers to their questions, particularly in areas such as DED [20].

According to the National Institute of Health, health-related content should generally be written at a reading level that is equivalent to or below the eighth grade[21]. The research findings suggest that the complexity of the texts produced by ChatGPT regarding DED corresponds to the expected understanding levels of those who have completed around 15 to 16 years of formal education. To control readability levels, artificial intelligence(AI) could be trained with updated parameters and material created under the guidance of a group of experts, which could help meet the necessary readability standards.

Materials on health information play a crucial role in patient education, allowing individuals to learn at their own pace, absorb information gradually, and share knowledge with essential people in their lives. Therefore, written health information documents' readability, comprehensibility, and quality are vital[22].

While ChatGPT offers instant responses about dry eye that are convenient and accessible for users' inquiries about DED, research suggests potential limitations regarding the quality of the content[20]. For example, Coşkun et al. reported that ChatGPT has difficulties in delivering precise and high-quality patient information on prostate cancer[23]. Cocci et al. found that ChatGPT produced low-quality information regarding urology patients [24]. Temel et al. reported that content created by chatGPT regarding spinal cord injury had substantial difficulties with quality [6]. Şahin et al. compared 5 different AI chatbots about erectile dysfunction and found that none of the chatbots had the required level of readability and quality[25]. Similar to all this research, our study also demonstrated that content about DED created by ChatGPT has significant quality issues. To address this issue, chatbots can be programmed to access trusted medical databases for producing health-related information, specific criteria can be created for the generated health content, and it can undergo evaluation by teams of specialists in the field.

This study has certain limitations that must be acknowledged. Limiting the analysis to only the initial 25 terms may have constrained the breadth of the investigation. Subsequent inquiries utilizing expanded keyword lists could yield a more exhaustive comprehension of the topic. Furthermore, the study exclusively concentrated on English keywords, which may restrict the applicability of the results. Incorporating queries in additional languages would enhance the breadth and diversity of the search results. The absence of content from other AI restricts the generalizability of the study's findings. Further investigation in this domain is essential to overcome these constraints and to improve our understanding of the subject matter.

CONCLUSION

In conclusion, this study revealed that the content generated by ChatGPT on DED exceeded readability standards and exhibited quality issues. To address these challenges, retraining the virtual intelligence with customized parameters and implementing content oversight by specialized teams and experts in content production could enhance its quality. More extensive research that encompasses a wider range of keywords, languages, and content from other AI sources within this field is needed.

Acknowledgements

Conflict of Interest

None declared

Data Availability Statement

The data that support the findings of this study are available from the corresponding author, (HIS), upon reasonable request.

Fundings

This study was not supported by any sponsor or funder.

Ethical Statement

As the study did not involve any operations conducted within a living organism, there was no need for permission from the ethical committee

Author Contributions: Conceptualization: HIS, BGS; Writing–Original Draft: HIS,BGS Writing–Review & Editing: HIS,BGS.

REFERENCES

1. The definition and classification of dry eye disease: report of the Definition and Classification Subcommittee of the International Dry Eye WorkShop (2007). *The ocular surface*. Apr 2007;5(2):75-92. doi:10.1016/s1542-0124(12)70081-2

2. Hopkins AM, Logan JM, Kichenadasse G, Sorich MJ. Artificial intelligence chatbots will revolutionize how cancer patients access information: ChatGPT represents a paradigm-shift. *JNCI Cancer Spectrum*. 2023;7(2)doi:10.1093/jncics/pkad010

3. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. MDPI; 2023:887.

4. Glick A, Taylor D, Valenza JA, Walji MFJJoDE. Assessing the content, presentation, and readability of dental informed consents. 2010;74(8):849-861.

5. Dave T, Athaluri SA, Singh SJFiAI. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. 2023;6:1169595.

6. Temel MH, Erden Y, Bağcıer FJWN. Information Quality and Readability: ChatGPT's Responses to the Most Common Questions About Spinal Cord Injury. 2024;181:e1138-e1144.

7. Erden Y, Temel MH, Bağcıer FJAoO. Artificial intelligence insights into osteoporosis: assessing ChatGPT's information quality and readability. 2024;19(1):17.

8. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. 2018;

9. Ladhar S, Koshman SL, Yang F, Turgeon RJCo. Evaluation of online written medication educational resources for people living with heart failure. 2022;4(10):858-865.

10. Moult B, Franck LS, Brady HJHe. Ensuring quality information for patients: development and preliminary validation of a new instrument to improve the quality of written health care

information. 2004;7(2):165-175.
11. Boles CD, Liu Y, November-Rider DJADHA. Readability levels of dental patient education brochures. 2016;90(1):28-34.
12. Vidal-Rohr M, Craig JP, Davies LN, Wolffsohn JS. The epidemiology of dry eye disease in the UK: The Aston dry eye study. *Contact lens & anterior eye : the journal of the British Contact Lens Association*. Jun 2023;46(3):101837. doi:10.1016/j.clae.2023.101837
13. McCann P, Abraham AG, Mukhopadhyay A, et al. Prevalence and Incidence of Dry Eye and Meibomian Gland Dysfunction in the United States: A Systematic Review and Meta-analysis. *JAMA ophthalmology*. Dec 1 2022;140(12):1181-1192. doi:10.1001/jamaophthalmol.2022.4394
14. Dana R, Bradley JL, Guerin A, et al. Estimated Prevalence and Incidence of Dry Eye Disease Based on Coding Analysis of a Large, All-age United States Health Care System. *American journal of ophthalmology*. Jun 2019;202:47-54. doi:10.1016/j.ajo.2019.01.026
15. Sonnenschein S, Stites ML, Gursoy H, Khorsandian JJES. Elementary-school students' use of digital devices at home to support learning pre-and post-COVID-19. 2023;13(2):117.
16. Stone DJJoD, Data, Practice DM. Mobile—more than a magic moment for marketers? 2012;13:280-294.
17. Shestak V, Slivinskaya N. Contemporary Approaches to Combat Cybercrimes in Ireland. 2021;
18. Sheppard AL, Wolffsohn JSJBoo. Digital eye strain: prevalence, measurement and amelioration. 2018;3(1):e000146.
19. Doak CC, Doak LG, Root JHJATAJoN. Teaching patients with low literacy skills. 1996;96(12):16M.
20. Ting DSJ, Tan TF, Ting DSWJE. ChatGPT in ophthalmology: the dawn of a new era? 2023:1-4.
21. Oliffe M, Thompson E, Johnston J, Freeman D, Bagga H, Wong PKJBo. Assessing the readability and patient comprehension of rheumatology medicine information sheets: a cross-sectional Health Literacy Study. 2019;9(2)
22. Bernier MJJON. Developing and evaluating printed education materials: a prescriptive model for quality. 1993;12(6):39-46.
23. Coskun B, Ocakoglu G, Yetemen M, Kaygisiz OJU. Can ChatGPT, an artificial intelligence language model, provide accurate and high-quality patient information on prostate cancer? 2023;180:35-58.
24. Cocci A, Pezzoli M, Lo Re M, et al. Quality of information and appropriateness of ChatGPT outputs for urology patients. 2023:1-6.
25. Şahin MF, Ateş H, Keleş A, et al. Responses of Five Different Artificial Intelligence Chatbots to the Top Searched Queries About Erectile Dysfunction: A Comparative Analysis. *Journal of medical systems*. Apr 3 2024;48(1):38. doi:10.1007/s10916-024-02056-0

Abbreviation:

AI: Artifical intelligence

GPT: Generative pre-training transformer

DED: Dry eye disease

EQIP: Ensuring Quality Information for Patients

FKGL: Flesch-Kincaid Grade Level

FKRE: Flesch-Kincaid Reading Ease

Figure Legends

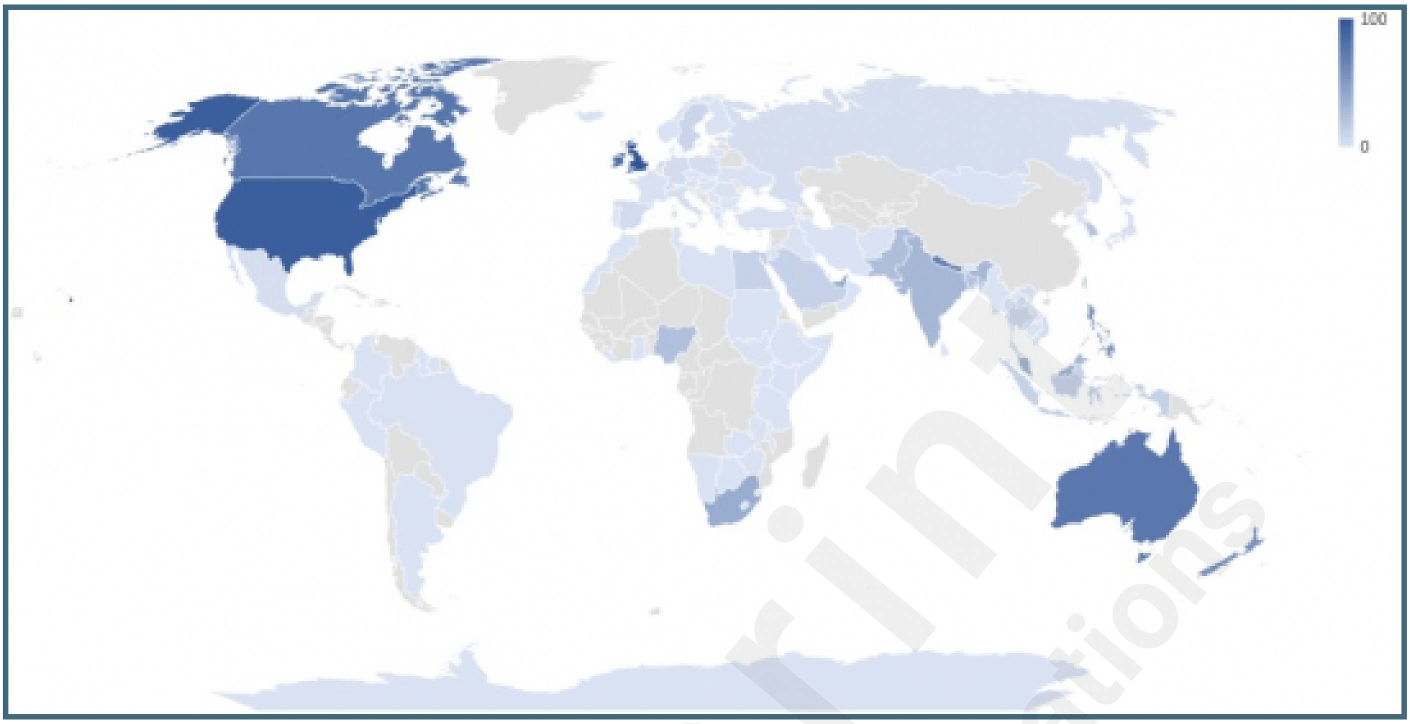
Fig *1*:The search interest in dry eye across different countries

Fig 2:Prevalence of dry eye from 2004 to present

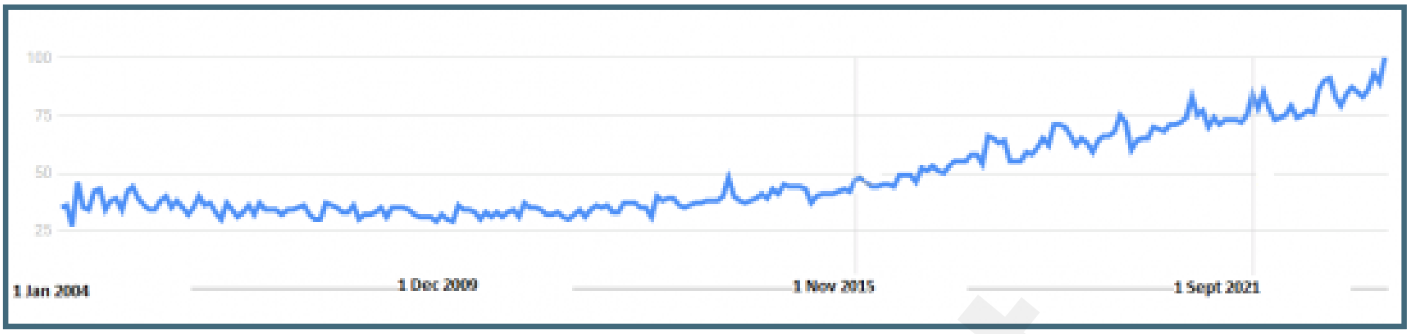
Supplementary Files

Figures

The search interest in dry eye across different countries.



Prevalence of dry eye from 2004 to present.


1 Jan 2004
1 Dec 2009
1 Nov 2015
1 Sept 2021