INTERNATIONAL NEUROUROLOGY JOURNAL

INJ

OPEN ACCESS

Insights Into Korean Public Perspectives on Urology: Online News Data Analytics Through Latent Dirichlet Allocation Topic Modeling

Young Wook Oh, JungYoon Kim

Department of Game Media, College of Future Industry, Gachon University, Seongnam, Korea

Purpose: The objective of the study was to explore how urology-related news, one of the medical specialties profoundly linked to human health and life, is communicated to the public through media outlets that serves as primary sources of medical information for the public.

Methods: In this study, articles were retrieved using the keyword 'Urology' from the Bigkinds spanning from January 1, 1990 to August 17, 2023. The Beautifulsoup library in Python was utilized for parsing the text to collect both titles and bodies of the articles. The gathered data was then analyzed using the latent Dirichlet allocation (LDA) algorithm from the scikit-learn library. Additionally, tools such as Wordcloud and Networkx were employed to visualize the relationships and patterns within the data.

Results: The keyword analysis led to the identification of various themes in the articles, with a clear distinction between those providing medical information and those promoting healthcare services, technologies, and products. Notably, the content frequently intertwined informational aspects with promotional ones. Articles on men's health and pet diseases, for example, often combined educational material with product or procedure promotions. This overlap highlights the complexity of categorizing media content into distinct themes. Furthermore, the coverage of health insurance and treatment methods including recent advancements like robotic surgery reflected the evolving nature of healthcare discussions to emphasize the interplay between policy changes, medical advancements, and media portrayal.

Conclusions: By identifying 10 distinct topics mentioned in the news, the analysis determined which topics are common in urology-related news coverage. The findings revealed a substantial volume of medical information on urology in the media with a wide range of topics including treatment and prevention of urologic conditions, insurance information, new treatments, and news stories promoting new products or hospitals.

Keywords: Urology; Online news; Public awareness; Data mining; Latent Dirichlet allocation

INTRODUCTION

Urology, as a critical domain within medical science, holds significant sway over human health and well-being. It covers a wide range of conditions that affect individuals across the population and exerts profound implications on the functionality and daily lives of those who are affected. While the field has witnessed substantial advancements in treatments and thera-

Corresponding author: JungYoon Kim https://orcid.org/0000-0002-2396-9514
Department of Game Media, College of Future Industry, Gachon University, 1342 Seongnamdaero, Sujeong-gu, Seongnam 13120, Korea
Email: kjyoon@gachon.ac.kr

peutic strategies, all underpinned by rigorous scientific inquiry, the dissemination of this specialized knowledge to the general public significantly lags behind.

Kim et al. demonstrated that traditional media outlets, specifically television and newspapers, have traditionally served as the primary conduits for conveying medical information to the public [1, 2]. However, the advent of digital media has reshaped the information landscape while elevating the Internet to a dominant platform for the dissemination and exchange of medical knowledge. It has surpassed traditional media in both reach and immediacy.

In this regard, this study attempts to unpack the shifts in public discourse surrounding urology by conducting a comprehensive analysis of journalistic content published for 27 years from 1996 through 2023. By examining this extensive collection of news articles, the study will gain a richer understanding of how public interest and media representation of urological issues have evolved, as well as the nature of the interplay between urological professionals and media outlets.

The advancements in computational technology have revolutionized our ability to perform textual analysis. What once required painstaking manual effort can now be accomplished with expediency and a greater degree of analytical precision thanks to the digitization of data and the computational power to process this big data. In particular, recent explorations into the use of big data for enhancing the understanding and treatment of interstitial cystitis/painful bladder syndrome underscore the innovative applications of big data in urological research [3]. The implications of big data for the medical field are vast and complex. While there is a clear potential to extract unprecedented insights from the accumulation of medical data, this must be carefully balanced against the ethical imperative to safeguard patient privacy and uphold the specialized nature of medical communication [4, 5].

Parallel to the rise of digital therapeutics is the burgeoning field of data-driven personalized medicine. Techniques like latent Dirichlet allocation (LDA), an algorithm conceived by Blei, Ng, and Jordan in 2003, offer a sophisticated mechanism to distill patterns and topics from expansive textual information. These methodologies are progressively charting the course for individualized patient care paradigms.

In pursuit of these insights, this study used LDA topic modeling to interrogate a dataset of 2,592 urology-related articles, with the dual objective of tracing the narrative thread of public concern as chronicled by the media, and showcasing the potential of advanced computational tools to enhance thematic exploration within medical journalism. The endeavor not only charts the historical narrative of urological discourse but also aims to forge pathways for enhancing the communication of medical breakthroughs to the public, thus bridging the knowledge gap identified at the outset.

The interplay between medical information and the news media presents a fertile area for research. News articles serve as critical conduits for conveying health information to the public and significantly shape perceptions and attitudes toward various health issues. Despite this pivotal role, the dissemination of health information through news has received comparatively little scholarly attention, especially when compared to fields such as politics or advertising. Yang and Park [6] provide valuable insights by classifying and analyzing the images in 330 health campaign articles from 2 local newspapers in Jeju Island. However, the study solely depends on researchers' personal categorical classification rather than statistical analysis, which restricts a comprehensive understanding of the wider landscape of media representations of health campaigns. While this method aids in suggesting future directions for health-related media content, it is still limited by a small sample size and a lack of statistical validation.

The advent of LDA topic modeling has fundamentally transformed text analysis, particularly in its ability to extract significant topics from extensive textual data. This technique is increasingly applied to analyze the content of news articles and scientific publications in the health sector. For example, Chang et al. [7] used LDA to identify public consumption and perception of nurses and highlight the need for strategic image-crafting by the nursing community. In similar way, Kim and Sun [8] used LDA to identify major topics in coronavirus disease 2019 (COVID-19)-related Korean academic papers to reveal 8 prominent topics.

Beyond the online new articles, Lee [9] focused on the consumption of medical information on YouTube by using Textstorm for data collection and LDA for analysis. It highlights the growing importance of digital platforms in public engagement with health information. Some research analyzed the consumption of COVID-19-related news on Twitter, a social network, in South Korea using the LDA methodology [10]. Additionally, this research examined the style or form of news coverage pertaining to COVID-19 in China [11, 12].

The previous literature shows that extensive and statistically robust analytical tools or methods can be utilized including LDA topic modeling. Then, a more detailed and empirically

grounded comprehension can be gained of how health information is disseminated through news media and absorbed by the public. In this regard, this study integrated LDA into the health communication to offer deeper insights into the thematic structures of health news content, thereby contributing to the development of more effective public health policies and media practices.

MATERIALS AND METHODS

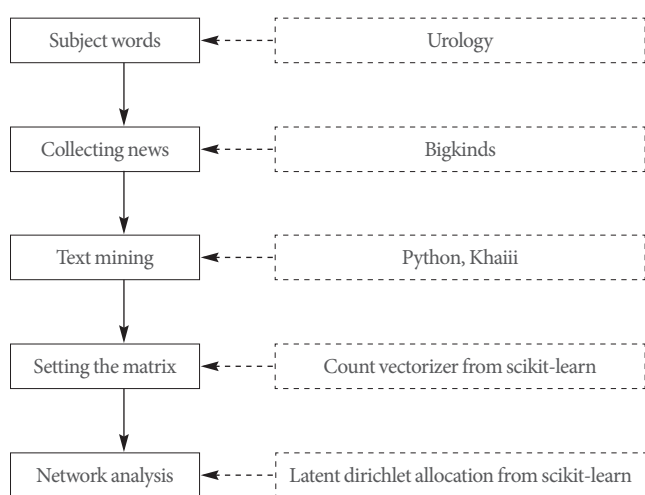The data collection and flow processing were handled as shown in (Fig. 1).

Fig. 1. Research process.

Data Collection

To assess the portrayal of urology in the media, a data retrieval was conducted by putting the search term 'urology' on BigKinds. The colleting news period covered from 1 January 1990 to 17 August 2023. We extracted a list of articles with their publication dates and URLs into an Excel file. Python's 'beautiful soup' library was used to parse the HTML content to capture article titles and bodies. From the initial 6,276 articles, the sample was reduced to 2,592 articles by removing spaces and inaccessible content. This curated set was then prepared for further analysis to understand trends and coverage of urology issues in the media.

Data Preprocessing

For data analysis and processing, a personal computer system was used. Despite the high computational demands of our tasks, we encountered no significant challenges when using a personal laptop. The computing setup comprises a MacBook Pro 16-inch, 2021 edition, with 32GB RAM and an Apple M1 Pro Silicon CPU. The computational tasks were principally conducted in the Python programming environment. As shown in Table 1, we opted for Python version 3.11 due to its robust stability and compatibility with the libraries required for our analyses. The suite of Python libraries applied for the LDA analysis is detailed subsequently in the methodology section.

The initial preprocessing step of the text data involved segmenting the text into morphological units using the Khaiii analyzer, which was crucial for subsequent analysis. To ensure the dataset's clarity and relevance, we eliminated surplus mor-

Table 1. Python development environment

Library	Version	Description
Sklearn (scikit-learn)	Scikit-learn = = 1.3.1 Sklearn = = 0.0.Post10	A library that provides simple and effective data mining and data analysis tools for use in Python.
Gensim	Gensim = = 4.3.2	A Python library for natural language processing, providing features such as topic modeling and document similarity analysis.
Khaiii	Build separately for osx	Korean morphology analyzer developed by Kakao
Pyldavis	Pyldavis = = 3.3.1	A library to provide latent Dirichlet allocation visualization tools in Jupyter Notebooks
Networkx	Networkx = = 3.2	A library to easily manipulate data with complex network structures in Python
Matplotlib	Matplotlib = = 3.8.0 Cycler = = 0.12.1 Kiwisolver = = 1.4.5 Pyparsing = = 3.1.1	Libraries for drawing graphs and visualizations in Python
Worldcloud	Wordcloud = = 1.9.2	Libraries used to visualize as word clouds
Os		For accessing files through the operating system from Python

phemes, such as unnecessary particles, adjectives, and idiomatic phrases. Initially, the focus in the study was on removing basic particles. However, it was recognized the necessity of expanding the list of exclusions to enhance the quality of our text analysis. This adjustment resulted from continuous data review including visualizations of word clouds and insights from the LDA analysis. The updated list of excluded terms signifies a balanced approach that filters out noise from the data while preserving the essential content required for LDA analysis.

LDA Topic Modeling

In this study, the refined text data underwent LDA topic modeling using the scikit-learn library in Python. Through LDA analysis, key topics and their associated significant words were identified. The parameters used in the extraction process are as follows (Table 2).

By setting the number of topics to 10 and the random seed to 0, we ensured that the same results could be reproduced. The process of updating the word distributions was iterated a maximum of 20 times. Both the document-topic prior and the topic-

Table 2. Latent Dirichlet allocation analysis parameter value

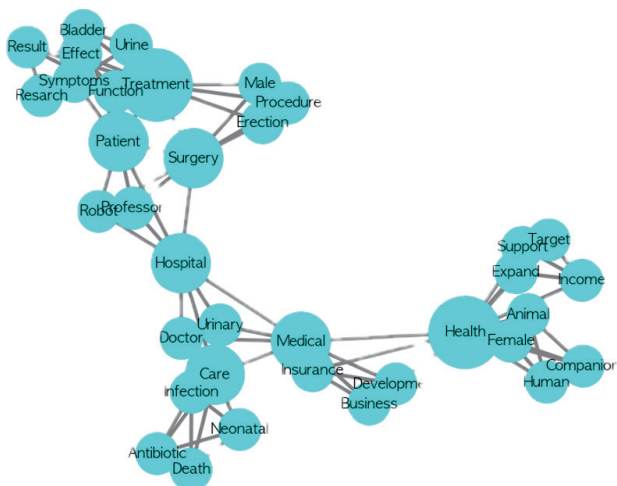| Parameter name	Value
n_components	10
random_state	0
max_iter	20
doc_topic_prior	0.1
topic_word_prior	0.1

word prior were set to 0.1, assuming a relatively even distribution of words across topics.

Organize Modeling Results

The extracted words were sorted according to their frequency, and the Python Library Networkx was used to construct a graph that represented the weight of the words that were shared between the different topics, as shown in (Fig. 2). This graphical representation made it easier to understand the relationships between the words.

The results show that 'treatment' emerges as a central focus across various topics that affirm its pivotal role in urology discussions. Simultaneously, 'health' and 'patient' appear consistently across several topics to indicate their widespread coverage and significance within urology-related contexts. Furthermore, terms like 'hospital,' 'surgery,' and 'medical' span multiple topics and it illustrates their interconnectedness and relevance to each other within the domain.

In order to highlight the lexical patterns within urology-related news discourse, a word cloud was constructed from the set of terms extracted from the articles. As shown in (Fig. 3), this graphical tool visually foregrounds the most common terms in order to provide a snapshot of the thematic focus of the field. It is interesting to note that 'treatment' and 'surgery' stand out prominently while underscoring their prevalence within urology articles. Moreover, the prominence of 'male' compared to 'female' suggests a higher frequency of medical information about men in urology-related content. This observation implies a potential disparity in the coverage of medical information, with a

Fig. 2. Word relationship.

Fig. 3. Word cloud.

Table 3. Ten topic with 5 key words produced by latent Dirichlet allocation and related news

No.	Topic	Word	Probability	Featured news
1	Neonatal deaths	Infection Neonatal Death Care Antibiotic	568.1928 412.5274 344.2218 293.6592 240.1809	"Citrobacter detected in 3 newborns at Ewha Womans University Medical Center"
2	Research and developments in new technologies and pharmaceuticals	Resarch Treatment Result Effect Function	1,154.2782 517.7516 459.2334 439.8429 432.4203	"Rituximab, a high-risk kidney transplant pretreatment drug, has no cancer-causing side effects with low dosage control" "< Korea's Top Hospital to Conquer Cancer > Wins 'Cancer' with Optimized Treatment System"
3	Urinary conditions that might relate to patients' everyday life	Treatment Urine Symptoms Bladder Patient	3,514.2926 2,400.9765 2,160.0514 2,131.1957 1,956.5907	"[Vivid Health 365] Severe flank pain and hematuria, 'Urolithiasis'" "Professionals Beware of Urolithiasis"
4	Broaden the scope of health screening initiatives	Support Target Expand Income Health	408.3253 343.1170 314.6315 289.5916 283.1545	"Kidney bladder anus ultrasound to be covered by health insurance starting in February" "Health insurance will cover ultrasounds of female genitalia, including fibroids, starting next year"
5	Pets and women's health	Female Health Human Animal Companion	566.2094 542.8198 465.1272 394.8786 312.3647	"Royal Canin Korea Launches 2021 Pet Weight Management Campaign" "'Home for the Holidays': Take Care of Your Fido's Health"
6	General medical terminology	Hospital Medical Doctor Medical Consultation Urology	1,183.6219 916.5285 783.7756 783.3318 696.4378	Skip
7	Accomplishments and public attention surrounding robotic surgery	Surgery Hospital Robot Hospital Patient	4,774.7334 1,881.6700 1,634.0999 1,324.8313 1,234.5908	"Bundang Cha Hospital Robotic Surgery...Over 500 cases in 2 years and 5 months" "[Healthy Family] "Robotic surgery, now a cost barrier, needs clinical guidelines to prevent abuse""
8	News about new product development and promotion	Health Insurance Medical Development Business	1,326.9323 1,157.1517 910.5208 616.3534 551.9146	"Newzy Labs Pharma subsidiary receives Phase 1 approval for diabetes combination drug, 'Expected to launch next year'" "[BioMedical] Jong-geun-dang, erectile dysfunction drug, 'Cialis generic war' begins"
9	Organs and organ donation	Uterus Patient Organ Disease Organ donation	436.1960 145.2952 123.7918 100.6522 89.1000	"Roundtable to promote organ and human tissue donation" "'Uterine fibroids' can be treated without a scalpel"
10	Erectile dysfunction in males	Male Surgery Erection Treatment Procedure	1,601.8156 1,405.6493 1,149.0781 1,006.7944 892.4635	"Did you take a pill while your wife was watching? You could always sneak them in." "Erectile dysfunction meds, pills taken to regain 'male pride' over sex" [M Interview].

focus on men's health receiving more attention than that of women within urology contexts.

RESULTS

To obtain optimal results, words were classified using Khaiii morphological analysis, and meaningless words that interfered with the analysis, such as investigations, adjectives, and idioms, were removed. Gossip articles related to celebrities were excluded. The results can be shown in (Table 3).

Of the 10 topics that emerged from the analysis, topics 1 and 6 were excluded from interpretation. Topic 1 was excluded from the analysis because of the severity of the situation and the high exposure of a single incident. Topic 6 is a collection of fairly generic words like "hospital" and "doctor," so it's hard to find a specific meaning. With the exception of topics 1 and 6, the interpretation of the remaining 8 topics is as follows.

Topic 2 includes the keywords 'research,' 'treatment,' 'result,' 'effect,' and 'function' and most of the articles were about research for the treatment of actual urinary diseases and the effect or function of new treatment methods. For examples, some articles were about the optimized treatment system or no side effects kidney transplant pretreatment drug. These articles often report research and results on new technologies or treatments.

Topic 3 includes the keywords 'treatment,' 'urine,' 'symptoms,' 'bladder,' and 'patient.' They focus on bladder health and related symptoms. In particular, bladder symptoms, such as urinary incontinence, are particularly common in the news because they are both common and uncomfortable in everyday life, and preventive measures tend to be covered in articles. It was also notable that it was covered equally in all years, not just the 2000s.

Topic 4 includes the keywords 'support,' 'aim,' 'expand,' 'income,' and 'health.' They are information about the availability of health insurance and healthcare coverage for urological conditions. Many news stories were informative about the expansion of insurance coverage for specific tests in a given year to provide news users with the information they needed. For example, the articles titled "Kidney Bladder Anus Ultrasound to be Covered by Health Insurance from February" and "Health Insurance to Cover Female Genital Ultrasound Examination for Uterine Fibroids and More from Next Year" were prominent in informing users that new tests had been included in health insurance coverage.

The main keywords in topic 5 are 'women,' 'health,' 'people,'

'animals,' and 'pets.' When analyzing these keywords, they can be interpreted as an expansion of services and products related to pet health. In particular, as pets have become more ingrained in people's lives, the number of services and products related to the treatment and management of pet diseases has increased, as has the number of news articles about their introduction.

Topic 7 includes the keywords 'surgery,' 'professor,' 'robot,' 'hospital,' and 'patient.' They discuss about innovations in medical technology, especially robotic surgery, which have increased in number since 2015. This indicates that robotic surgery is beginning to become popular as hospitals that have introduced robots advertise their achievements.

In topic 8, the keywords such as 'health,' 'insurance,' 'medical,' 'development,' and 'business' focus on development and business rather than medical keywords. There were many articles about the development and insurance coverage of new technologies, and news about the business performance and expectations of new treatment technologies or drugs for urinary diseases. "Newzy Labs Pharma Subsidiary Approves Diabetes Combination Drug in Phase 1 'Expected to Launch Next Year'" and "[BioPharma] Zhongroxin, Erectile Dysfunction Treatment, 'Cialis Generic War' Begins" were among the many articles in which companies announced their development and expected economic impact.

Topic 9 includes the keywords 'uterus,' 'patient,' 'organ,' 'disease,' and 'organ donation.' Numerous articles centered on exploring the effects of the uterus on diseases related to the urinary system. The inclusion of the keyword 'organ donation' may be attributed to a urology professor's involvement in organ donation symposiums, particularly emphasizing kidney transplantation, a major surgery associated with organ donation. For instance, some articles highlighted a urologist's participation in kidney transplant procedures and another article introduced a condition associated with the uterus, potentially related to the discussions within this theme.

Topic 10 includes the keywords 'dysfunction,' 'surgery,' 'erection,' 'male,' 'treatment.' Like the topic 4, the keywords showed that male sexual dysfunction was of particular interest to the public. In particular, articles on erectile dysfunction appeared throughout the ages, and different types of articles appeared, from columns to new drug development and advertising. Several articles featured introductions to new drugs designed for addressing erectile dysfunction, accompanied by interviews with doctors providing explanations advocating for their usage.

DISCUSSION

In categorizing the 8 topics, it becomes evident that discussions regarding new drugs and technologies in urology swiftly reach the public as exemplified by topic 2 (research and developments in new technologies and pharmaceuticals) and topic 8 (new product development and promotion). Second, news about robotic surgery (topic 7), and specialized information on surgical treatments like kidney transplantation (topic 9), are also widely disseminated. Thirdly, an array of urological diseases is introduced including from everyday urinary conditions (topic 3) to women's health (topic 5), male sexual dysfunction (topic 10), and pet-related health matters. Lastly, information concerning urology diseases covered by health insurance (topic 4) undergoes annual updates and is effectively communicated to the public.

A significant number of articles focused on the promotion of hospitals, doctors and products. However, there were also a significant number of informative articles about urological conditions. Thus, the 2 types of article often intermixed. Notably, articles on men's health, particularly in topic 10, were closely associated with everyday life to feature a range of products, treatments, and at times, traditional remedies. These articles exhibited consistent appearances from 1990 to 2023.

Simultaneously, information regarding the expansion of insurance coverage, observed in topics 8 and 4, was prevalent. This suggests that discussions about changes in insurance policies often occurred alongside promotions for specific medical products or hospitals. This indicates a close connection between health insurance utilization and the broadening of treatment options. The emergence of 'robotic surgery' as a keyword in topic 7 since 2011 indicates that the practice of robotic-assisted medical procedures is relatively recent and not long established. Recent news stories have also highlighted the financial constraints associated with this technology.

This study investigated trends in urology-related news by analyzing news articles retrieved by searching for urology-related terms from 1990 to 2023. Through this method, 8 meaningful distinct themes were analyzed in urology-related news. As evidenced in the analysis, there is a substantial dissemination of medical information concerning urology across various media platforms. This information encompasses details regarding the treatment and prevention of urological diseases, insights into insurance aspects, updates on new treatments, and coverage of developments related to new products and hospitals.

In this study, one limitation is the lack of detailed exploration into each individual article. Focusing solely on the broader term 'urology' overlooked specific nuances associated with conditions like urinary incontinence or prostate disease, which are crucial for public perception and understanding. Therefore, future studies can be done to enhance their depth by targeting more specific keywords related to particular diseases within urology. This will uncover prevalent information and gain a better understanding of the public's perception of these specific conditions by providing a more comprehensive view of how these diseases are portrayed and understood in the public sphere.

AUTHOR CONTRIBUTION STATEMENT

· Conceptualization: *JYK*
· Data curation: *YWO*
· Formal analysis: *YWO*
· Funding acquisition: *JYK*
· Methodology: *YWO, JYK*
· Project administration: *JYK*
· Visualization: *YWO, JYK*
· Writing - original draft: *YWO*
· Writing - review & editing: *JYK*

ORCID

Young Wook Oh   0009-0009-7862-2655
JungYoon Kim   0000-0002-2396-9514

REFERENCES

1. Kim DY, Kwon OD, Kim SG, Shin IH. Recognitional study about patients and caregivers' understanding of clinical trial. J Korean Cont Soc 2008;12:48-57.
2. Choi H, Park JY, Yeo JK, Oh MM, Moon DG, Lee JG, et al. Population-based survey on disease insight, quality of life, and health-seeking behavior associated with female urinary incontinence. Int Neurourol J 2015;19:39-46.
3. Cho YW, Youn SH. Intravesical bladder treatment and deep learning applications to improve irritative voiding symptoms caused by interstitial cystitis: a literature review. Int Neurourol J 2023;27:13-20.
4. Kim JY. Era of the fourth industrial revolution and the urologists' journey to navigating big omics data. Int Neurourol J 2018;22:101-2.

5. Kim EJ, Kim JY. The metaverse for healthcare: trends, applications, and future directions of digital therapeutics for urology. Int Neurourol J 2023:27:3-12.

6. Yang NM, Park KS. An analysis of medical information and health campaign contents reported in local daily newspapers. Korean J Soc Theory 2021:60:139-75.

7. Chang SJ, Park SA, Son YD. A topic modeling analysis for Korean online newspapers: focusing on the social perceptions of nurses during the COVID-19 epidemic period. J Korean Acad Soc Nurs Educ 2022:28:444-55.

8. Kim EH, Sun YH. Research topic analysis of the domestic papers related to COVID-19 using LDA. J Korea Inst Inf Electron Commun Technol 2022:15:423-32.

9. Lee SJ. An analysis on major issues of medical information on You-Tube platform using LDA topic modeling. J Korea Content Assoc 2023:23:198-207.

10. Park HW, Park S, Chong M. Conversations and medical news frames on Twitter: infodemiological study on COVID-19 in South Korea. J Med Internet Res 2020:22:18897.

11. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health communication through news media during the early stage of the CO-VID-19 outbreak in China: digital topic modeling approach. J Med Internet Res 2020:22:19118.

12. Culha Y, Culha MG, Acaroglu R. Evaluation of YouTube videos regarding clean intermittent catheterization application. Int Neurourol J 2020:24:286-92.