SYSTEMATIC REVIEW

Analysis of Publication Activity and Research Trends in the Field of Lichen Planus: Pubmed Review

Aida Khakimova¹* and Oleg Zolotarev¹

¹Institute of Information Systems and Computer Engineering, Russian New University, 105005, Moscow, Russia

Abstract:

Background:
Currently, there is only one bibliometric study of lichen planus (LP) and oral lichen planus (OLP) in the literature, which examined the most cited articles in the Scopus database. Our study covered all published articles in the PubMed database for 140 years since, from 1880 to 2021. In addition to the classical bibliometric analysis, we conducted a lexical analysis of key terms to build research trends in the oral lichen planus.

Aims:
Analysis of publication activity in the field of lichen planus by countries, their economic status and population, as well as identification of concomitant diseases by lexical analysis of key terms extracted from headings and abstracts over the past 20 years.

Methods:
Information from the PubMed database was retrieved automatically based on a query for the period from 1880 to 2021, including the field: title, abstract, authors, and year of publication. A total of 8173 articles were retrieved. The selection of keywords and identifying trends in related terminology were done using expert and automatic methods.

An analysis of publication activity by country and socio-economic indicators was carried out. With the help of neural network analysis, the most characteristic terms related to LP were identified. Common terms were ranked by occurrence in titles and abstracts.

Results:
Publication activity in the field of LP and OLP has especially increased in the 21st century. The United States is the most productive country. China is a leader among countries with economies in transition. India is a leader among emerging economies. LP research is distributed worldwide. Finland ranks first in the number of publications per capita. The temporal dynamics of terminology are noted, including an increase in the number of terms used in any field of science (hereinafter referred to as general scientific terms).

Conclusion:
Publication activity in the field of LP and OLP has increased significantly in the 21st century. The highest publication activity was observed among authors from India, the USA and China. The leaders among economically developed countries are Italy, among the countries with economies in transition - China, and among the emerging economies - India. Based on lexical analysis of key terms, the following concomitant diseases were identified: carcinoma, leukoplakia, hepatitis, and lupus.

Keywords: Lichen planus, Oral lichen planus, Publication activity, Research trends, Key terms, Bibliometrics.

1. INTRODUCTION

This study analyzes the researchers’ performance in the field of lichen planus (LP). The data were extracted from the PubMed database with the search parameters: relative growth rate, country-wise distribution, and prevailing vocabulary.

Lichen planus is a chronic inflammatory, immunodependent disease of the skin and mucous membranes [1]. In general structure of dermatological morbidity, LP is 1.5-2.4%, among all diseases of the oral mucosa — 30-35% [2]. Despite the prevalence, only one bibliometric analysis
related to LP in the academic literature has been carried out.

Usually, bibliometrics studies focus on a single combination of words. Our approach is based on the study of terms using contextual analysis [3].

2. SUBJECTS AND METHODS

PubMed is MEDLINE’s public search engine for biomedical literature [4]. PubMed has become a very popular resource among scientists interested in research publications in various biomedical fields [5]. One of the most valuable features of PubMed is users’ ability to restrict search through different settings. Users can target their search by restricting the query to publications that meet various requirements (including the time of publication, type of manuscript, type of research, language, authors’ affiliation, journal, etc.) [6].

Medical Subject Headings (MeSH) for scientometric analysis in the medical context have increased in the last few years. The MeSH terms have been used for identifying hot topics and research areas to map the dynamics of emerging technologies based on co-word analysis [7, 8].

The data for our study were extracted from the PubMed database. The following search keywords were used (search query fragment): “lichen planus” OR “lichen planus pigmentosus” OR “hypertrophic lichen planus” OR “mucosal lichen planus” OR “annular lichen planus” OR “hypertrophic lichen planus” OR “atrophic lichen planus” OR “ulcerative lichen planus” OR “bullous lichen planus” OR “lichen planus pemphigoid” OR “inverse lichen planus” OR “actinic lichen planus.”

We analyzed articles published from 1880 to 2021. A total of 8173 records for LP and 4775 records for OLP were extracted from the PubMed database for this study (15.01.2021).

The papers from different countries were identified by using the Advanced Search Builder. The number of publications by country was determined by inserting the name/s of the countries in the «affiliation» field in the «advanced» section of the PubMed search engine, for example, (“lichen planus”(MeSH Terms) OR to (“lichen”(All Fields) AND “planus”(All Fields)) OR “lichen planus”(All Fields)) AND Andorra(Affiliation)), (“lichen planus”(MeSH Terms) OR (“lichen”(All Fields) AND “planus”(All Fields)) OR “lichen planus”(All Fields)) AND USA(Affiliation) OR United States(Affiliation). The search query for the oral lichen planus was the following: (“lichen planus, oral”(MeSH Terms) OR (“lichen”(All Fields) AND “planus”(All Fields)) OR “oral”(All Fields)) OR “oral lichen planus”(All Fields) OR “oral”(All Fields) AND “lichen”(All Fields) AND “planus”(All Fields)) AND (“0001/01/01”(PubMedDate): “2020/12/31”(PubMedDate)).

To calculate the number of publications per country, we searched for publications by authors’ affiliation. Then we summarized the number of publications with known affiliation and expressed the number by country as a percentage of the total number of extracted publications. To find the number of publications per million persons or gross domestic product (GDP), we divided the number of publications per country’s population (or per GDP).

A special search strategy was developed for the present study. To extract keywords from the titles and abstracts, we used MEDLINE/PubMed database for all years. A set of programs in the python language was developed. The programs extract articles with certain keywords/terms in titles and abstracts and perform statistical terms’ analysis of the MEDLINE/PubMed database records (titles, abstracts, authors, etc).

Based on the terminology statistics of thematic collections, we determined the words and collocations more often found together with the studied terms (the key terms of collection, for example, “lichen planus”) in titles and abstracts by periods.

From the entire database, we identified articles with the term “lichen planus,” breaking them down over the years. The process includes two steps. The first step uses a semantic selection of appropriate terms with a short context size (with the 3 words window). Next, semantic information is used to select the prevalent terms by year.

For further analysis, we chose three periods: 1960-1979s, 1980-1999s, and 2000-2020s. Therefore, we divided all terms into three groups. Each group was processed statistically, highlighting the main keywords and terms by occurrence frequency. Since the periods vary greatly in the total number of publications, the approach was as follows. For the period of 1960-1979, we selected words that occur at least 2 times a year for analysis. For the period of 1980-1999, we selected words occurring at least 3 times a year. For the period of 2000-2020, we selected words occurring at least 5 times a year. If the terms occurred less than five times in the year, they were ignored. In the resulting set of terms, thematic groups were identified, and the occurrence frequency of terms was analyzed.

When researching the trends of concomitant diseases, the search was built according to the scheme: (“lichen planus”(MeSH Terms) OR (“lichen”(All Fields) AND “planus”(All Fields)) OR “lichen planus”(All Fields)) AND (“leukoplakia”(MeSH Terms) OR “leukoplakia”(All Fields)).

Example of the Search query: (LICHEN PLANUS) AND LUPUS included to (“lichen planus”(MeSH Terms) OR (“lichen”(All Fields) AND “planus”(All Fields)) OR “lichen planus”(All Fields)) AND (“lupus vulgaris”(MeSH Terms) OR (“lupus”(All Fields) AND “vulgaris”(All Fields)) OR “lupus vulgaris”(All Fields) OR “lupus”(All Fields) OR “lupus erythematosus, systemic”(MeSH Terms) OR (“lupus”(All Fields) AND “erythematosus”(All Fields) AND “systemic”(All Fields)) OR “systemic lupus erythematosus”(All Fields)).

The research flow chart is presented in (Fig. 1).
Analysis of trends was carried out using our programs.

3. RESULTS

3.1. Publications by Country and Globally

The search was carried out in January 2021 in PubMed Central, limited to scientific articles in English as the main language.

A sharp increase in publication activity began in the 80s of the last century. Trend analysis for every year showed exponential growth ($y = 1.5641e^{0.0507x}$, $R^2 = 0.8388$ for LP, $y = 0.791e^{0.0809x}$, $R^2 = 0.9483$ for OLP).

We conducted a comparative analysis of publication activity over decades. 62.00% of the total publications were published in the 21st century (Fig. 2).

![Flow chart](image1.png)

**Fig. (1).** The research flow chart.

![Graph](image2.png)

**Fig. (2).** The percentage of publications related to LP and OLP by decades.
We determined the number of publications by the country for the total period. We divided the sum of publications for all countries by the total number of world publications (obtained without searching any country name in the “Affiliation” field, see Methods section).

We defined the most productive 20 countries in LP research from 1880 to 2021 (current data, 15.01.2021). India stands out in first place with 1211 papers (17.11% of whole world publications), followed by the United States with 834 publications (11.79%), China with 556 (7.86%), Italy with 444 (6.27%), and Iran with 328 (4.64%). 74 countries such as Albania, Andorra, Angola, Antigua and Barbuda, Azerbaijan, Bahamas, Belarus, Belize, Bhutan, Bolivia, Botswana, et al. have no publications yet.

3.2. Publications Relative to Population and GDP

We found the number of publications per million persons (PMP) [9]. Barbados (34.84) and Switzerland (9.87) ranked first and second, followed by Croatia with 9.42 publications per million persons (Fig. 3). Uganda (0.02) and the Democratic Republic of Congo (0.01) ranked last.
We also investigated whether the Gross Domestic Product (GDP) influences LP research productivity in all countries. As for bias regarding the GDP of each country, it was accounted for by finding the number of LP-related publications per GDP (in US Billion Dollars by International Monetary Fund (2020)). We reviewed the ratio of LP publications per capita nominal GDP [10]. We divided countries into groups according to the economic classification of the International Monetary Fund (economically developed countries, countries with economies in transition, and developing countries) [10, 11].

Fig. (4) shows economically developed countries. The undisputed leaders are Italy, the USA, the UK, Spain, Japan, Germany, and Greece.

Among the countries with economies in transition, China is the undisputed leader, and Poland and Georgia are leaders, too (Fig. 5).

Among the emerging economies, India is the undisputed leader. Iran and Brazil also stand out (Fig. 6).

**Fig. (5).** Ratio of publications to GDP per capita for countries with economies in transition. The order of countries corresponds to an increase in GDP per capita.

**Fig. (6).** Ratio of publications to GDP per capita for developing countries. The order of countries corresponds to an increase in GDP per capita.
3.3. Lexical Analysis of Key Terms

We carried out a statistical analysis of the headings and abstracts of articles related to the disease. The most characteristic phrases, including the term “lichen planus,” were extracted from the database. Moreover, the analysis of the semantic environment of key terms allows for representing the research trends. Terms related to LP were highlighted. We processed the selected terms for the periods 1960-1979, 1980-1999, and 2000-2020. After the exclusion of rare terms (see the methods section), the remaining terms were divided into lexical and semantic groups: “Medical terms related to disease,” “General medical terms,” “General scientific terms,” and “Other vocabulary.” The terms ranged within the group.

An analysis showed that the ratio of terms in titles changes over time (Fig. 7).

Over time, general medical and scientific terms are more often used in the titles. Table 1 shows the most common terms for lexical and semantic groups.

3.4. Trend Analysis by Key Terms

By specific terms, it is possible to determine the actual areas of research. For 60 years, the term “oral” has prevailed; therefore, the lesion of the oral mucosa (mucosa) is considered. Skin and nails are also studied. In 1960-1979, “ruber” (“Lichen ruber planus”) was often used. In the next 20 years, the use of this combination of terms decreased. In the last 21 years, the term “lichen planus” has been used more often.

Studies increasingly mention the cellular level, considering various cells (squamous cell, cell, carcinoma, and t-cells). The term “erosive” is constantly used as a characteristic. In the 2000s, the term “lichenoid” was often used.

From 1960-1979, researchers examined lichens along with lupus. From 1980-1999, they began to study the relationship between lichen with hepatitis and viral infection. In the 2000s, hepatitis research was ongoing, and the oncological aspect was being studied - the relationship between carcinoma and malignant diseases.

![Fig. (7). Lexico-semantic groups of keywords.](image)

<table>
<thead>
<tr>
<th></th>
<th>Specific Medical Terms (10 Prevalent), %</th>
<th>General Medical Terms (5 Prevalent), %</th>
<th>Scientific Terms (5 Prevalent), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960-1979</td>
<td>oral, 30.5; ruber, 27; mucosa, 10.25; lesions, 7,75; cells, 7; skin, 5,5; nail, 3,5; lupus 3,25; erosive 2,75; mucous 2,5</td>
<td>treatment 41,88; patients 21,37; clinical 15,38; therapy 11,97; disease 9,40</td>
<td>cases 32,82; study 40,46; acid 12,21; vitamin 7,63; cavity 6,87</td>
</tr>
<tr>
<td>1980-1999</td>
<td>oral 28,47; cells 7,85; mucosa 6,66; lesions 6,02; hepatitis 5,77; ruber 5,26; chronic 4,62; skin 4,33; erosive 3,56; virus 3,35</td>
<td>patient(s) 33,42; treatment 20,35; disease 18,46; clinical 16,44; therapy 6,20</td>
<td>association(ed) 37,97; study 19,33; case 15,99; expression 9,32; report 5,87</td>
</tr>
<tr>
<td>2000-2020</td>
<td>oral 36,93; cell 6,08; lesion(s) 5,51; hepatitis 5,07; erosive 4,57; malignant 3,54; lichenoid 3,54; carcinoma 2,86; skin 2,75; virus 2,58</td>
<td>patient(s) 33,67; treatment 18,17; clinical 14,25; disease(s) 13,72; therapy 6,15</td>
<td>case 14,75; study 14,42; association (ed) 9,67; review 9,15; expression 7,55</td>
</tr>
</tbody>
</table>
To confirm the findings, we reviewed PubMed's statistics on concomitant diseases - carcinoma, leukoplakia, hepatitis, and lupus (search in PubMed).

Oral LP is known to be related to chronic liver diseases. Hepatitis is a liver dysfunction with an impaired immune system and is considered a potential disease that may increase the risk of OLP. It was shown that in comparison to the general population, the risk of developing OLP in patients infected with the Hepatitis C virus is twice higher [12].

The presence of Candida albicans aggravates the course of chronic inflammatory diseases (OLP, lupus erythematosus, leukoplakia). OLP and leukoplakia are two precancerous lesions of great importance in the pathology of the oral cavity [13]. Lichen planus and lupus erythematosus rarely coexist in the same patients [14].

Most publications link LP and carcinoma. The trend is polynomial: \( R^2 = 0.8118 \) (Fig. 8a). The trend of joint description of OLP and carcinoma is exponential: \( y = 1.0312e^{0.0636x}, R^2 = 0.8955 \) (Fig. 8b).

Fig. (8). Trends in publication activity in the area of lichen planus and related diseases (carcinoma, leukoplakia) (A); of oral lichen planus and related diseases (carcinoma, leukoplakia) (B).
A polynomial trend character is a characteristic of joint description of LP and leukoplakia: \( R^2 = 0.6919 \) \( (y = 0.0078x^2 - 0.0594x + 4.7188) \) (Fig. 8a). There is the same character of the trend of joint description of OLP and leukoplakia: \( y = 0.0039x^2 + 0.2284x, R^2 = 0.6364 \) (Fig. 8b).

Analysis of publication activity trends in lichen planus and related diseases (lupus, hepatitis) Fig. (9) showed:

- fuzzy linear trend for LP and hepatitis \( (y = 0.3251x, R^2 = 0.5426) \) with a surge in activity in 1995-2007; polynomial trend with a plateau for OLP and hepatitis \( (y = -0.0027x^2 + 0.3272x, R^2 = 0.2405) \) (not shown);
- fuzzy polynomial trend for LP and lupus \( (y = 0.0056x^2 - 0.0896x + 5.6223, R^2 = 0.453) \); exponential trend for OLP and lupus \( (y = 1.3214e^{0.0255x}, R^2 = 0.4219) \) (not shown).

5. DISCUSSION

The first mention of the term “Lichen planus” was noted in 1880. A graphical representation of publication activity revealed the exponential nature of the trend. The magnitude of the accuracy of the approximation \( (R^2) \) is 0.8388. \( R^2 \) is close to 1, so the selected function accurately approximates the data. A comparative analysis of publication activity over the decades showed that more than 62% of the total publications were published in less than 20 years of the 21st century.

Of the most productive 20 countries in LP research, India stands out in the first place with 17.11% of the total number of publications, followed by the USA with 11.79%, China with 7.86%, Italy with 6.27%, and Iran with 4.64%. Of these countries, four are located in Asia, nine in Europe, three in Near East, two in North America, one in South America, and one in Oceania. So LP is a topic of interest in several regions of the world. LP has a worldwide distribution with no overt racial predisposition.

We distinguish groups of countries according to the number of publications:

- more than 300 – India, United States of America, China, Italy, Iran,
- from 100 to 300 – Brazil, United Kingdom, Japan, Germany, Turkey, Spain, Poland, Canada, Australia, France, Saudi Arabia,
- from 50 to 100 – Taiwan, Romania, Switzerland, Greece, Thailand, Israel, Austria, Belgium, Egypt,
- from 10 to 50 – Finland, Sweden, Portugal, Malaysia, Croatia, Georgia, Mexico, Denmark, Pakistan, South Africa, Singapore, Norway, Hong Kong, Chile, Lebanon, Argentina, Russia, Hungary, Nigeria, Colombia, Ireland, Nepal, Morocco, Iraq, South Korea, Tunisia, Kuwait, New Zealand, Serbia, Sri Lanka, Qatar,
- less than 10 – Bulgaria, Jordan, Indonesia, Sudan, Benin, Oman, Syria, United Arab Emirates, Vietnam, Bangladesh, Bosnia and Herzegovina, Estonia, Tanzania, Yemen, Côte d’Ivoire, Ethiopia, Ghana, Guatemala, Kenya, Mali, Slovenia, Venezuela, Cameroon, Libya, Malta, Philippines, Uruguay etc.

The absolute number of scientific publications does not fully reflect the «scientific situations» in the country since it strongly depends on the size of the population. Barbados ranked first with 34,84 publications per million persons, followed by Switzerland (9,87) and Croatia (9,42).

We also investigated whether the Gross Domestic Product (GDP) influences LP research productivity in all countries by economic groups according to the classification of the International Monetary Fund.
The undisputed leaders among the economically developed countries are Italy, the USA, the UK, Spain, Japan, Germany, and Greece, the countries that entered the top 20 in the total number of publications. Among the countries with economies in transition, China is the undisputed leader. Among the emerging economies, the leaders are India, Iran and Brazil.

In general, we can conclude that not all developed countries conduct research in the field of dermatology. Therefore, the Czech Republic and Iceland had a single publication, 2 articles are from Slovakia and Cyprus, 3 are from Malta, and 4 are from Slovenia.

Countries with economies in transition and countries of the former USSR either do not have publications at all or have very few. There is no publication from Turkmenistan, Belarus, Moldova, Azerbaijan, Tajikistan, or Kyrgyzstan. There are few publications from Kazakhstan, Armenia, Uzbekistan (1 publication); Latvia, Lithuania, Ukraine (2 publications); Estonia (5 publications), Russia (21 publications), and Georgia (33 publications). Mongolia, Albania, and Laos also have no publications. There are few publications from Cambodia (1), North Macedonia (2), Bosnia and Herzegovina (5), and Vietnam (6 publications).

Our approach permits more detailed information about the terms, their characteristics, related factors, and processes. A statistical analysis of the headings and abstracts of articles related to LP made it possible to distinguish the following lexical and semantic groups: “Medical terms related to disease,” “General medical terms,” “General scientific terms,” and “Other vocabulary.” We analyzed the highlighted groups of terms for the periods 1960-1979, 1980-1999, and 2000-2020 (see the methods section). The total number of words selected in 1960-1979 was 664, in 1980-1999 - 4413, 2000-2020 - 11168.

A change in the ratio of terms in titles over time has been found. In each reviewed period, the titles were dominated by specific medical terms (related to the disease). In each period, general scientific terms were in the second place in use, and general medical terms were in the third place. However, the dynamics of the correlation of terms are noted. From the period 1960-1979 to the modern period, there is a clear tendency to reduce the frequency of use of specific terms (by 8.86%), while the number of general medical (by 2.48%) and general scientific terms, especially the latter (by 7.78%), is growing. There are examples of general scientific terms: association, disease, “Other vocabulary.” We analyzed the highlighted groups of terms for the periods 1960-1979, 1980-1999, and 2000-2020 (see the methods section). The total number of words selected in 1960-1979 was 664, in 1980-1999 - 4413, 2000-2020 - 11168.

A change in the ratio of terms in titles over time has been found. In each reviewed period, the titles were dominated by specific medical terms (related to the disease). In each period, general scientific terms were in the second place in use, and general medical terms were in the third place. However, the dynamics of the correlation of terms are noted. From the period 1960-1979 to the modern period, there is a clear tendency to reduce the frequency of use of specific terms (by 8.86%), while the number of general medical (by 2.48%) and general scientific terms, especially the latter (by 7.78%), is growing. There are examples of general scientific terms: association, disease, “Other vocabulary.” We analyzed the highlighted groups of terms for the periods 1960-1979, 1980-1999, and 2000-2020 (see the methods section). The total number of words selected in 1960-1979 was 664, in 1980-1999 - 4413, 2000-2020 - 11168.

A change in the ratio of terms in titles over time has been found. In each reviewed period, the titles were dominated by specific medical terms (related to the disease). In each period, general scientific terms were in the second place in use, and general medical terms were in the third place. However, the dynamics of the correlation of terms are noted. From the period 1960-1979 to the modern period, there is a clear tendency to reduce the frequency of use of specific terms (by 8.86%), while the number of general medical (by 2.48%) and general scientific terms, especially the latter (by 7.78%), is growing. There are examples of general scientific terms: association, disease, “Other vocabulary.” We analyzed the highlighted groups of terms for the periods 1960-1979, 1980-1999, and 2000-2020 (see the methods section). The total number of words selected in 1960-1979 was 664, in 1980-1999 - 4413, 2000-2020 - 11168.

A change in the ratio of terms in titles over time has been found. In each reviewed period, the titles were dominated by specific medical terms (related to the disease). In each period, general scientific terms were in the second place in use, and general medical terms were in the third place. However, the dynamics of the correlation of terms are noted. From the period 1960-1979 to the modern period, there is a clear tendency to reduce the frequency of use of specific terms (by 8.86%), while the number of general medical (by 2.48%) and general scientific terms, especially the latter (by 7.78%), is growing. There are examples of general scientific terms: association, disease, “Other vocabulary.” We analyzed the highlighted groups of terms for the periods 1960-1979, 1980-1999, and 2000-2020 (see the methods section). The total number of words selected in 1960-1979 was 664, in 1980-1999 - 4413, 2000-2020 - 11168.


