

# Trends in scientific output on the lesbian, gay, bisexual, and transgender (LGBT) community research: A bibliometric analysis of the literature

David AN<sup>1,\*</sup>, Kavita BATRA<sup>2</sup>

## Affiliations:

<sup>1</sup> *Culver Academies, 1300 Academy Road, Culver, Indiana 46511, USA. ORCID: 0000-0002-3424-2740.*

<sup>2</sup> *Department of Medical Education/Trauma & Critical Care, Kirk Kerkorian School of Medicine, University of Nevada, Las Vegas, Nevada, United States of America. E-mail: kavita.batra@unlv.edu ORCID: 0000-0002-0722-0191.*

\*Corresponding Author:

*David An. Academy Road, Culver, Indiana 46511, USA. Email: anculver2023@gmail.com; david.an@culver.org*

## Abstract

**Introduction:** Lesbian, Gay, Bisexual, and Transgender (LGBT) represent a diverse group with special needs due to the unusual developmental experiences and social inequalities. This paper aims to explore and outline a future research direction in LGBT issues through tracing our historical understanding of this population from an aspect of scientific research.

**Methods:** LGBT-related peer-reviewed documents were retrieved from the PubMed database and the study period was set from the inception to 2021. Python-based methods were then performed to analyze the publication metadata and extract the most prominent research topics based on the abstract contents. Key points covered in the study were the development and trend of scientific effort and research themes in the LGBT topic, identified through the Bigram model and Latent Dirichlet Allocation algorithm.

**Results:** A total of 21,221 publication records were retrieved from the PubMed database. Literature analyses demonstrated that scientific research in LGBT had grown gradually but began to gain momentum since 2010, evidencing increased attention to this demographic in the last decade. Regarding the region-wise scientific effort in LGBT, the United States (U.S.) was the most productive country (with > 45% of the total publications), followed by the United Kingdom (UK), Canada, Australia, and the Netherlands. Furthermore, Peru and Thailand, besides the U.S., Australia, and Canada, were the top countries that had relatively allocated more of their scientific efforts to LGBT research based on the calculated activity indices. Topics attracting the most attention in LGBT research over time were "male sexuality and risk", followed by "sexual development", "health care service", "social experience", and "intervention strategies".

**Discussion:** This study provided a broad view of the developmental trends in LGBT research from invisibility to attention through a bibliometric lens and could serve as a data-based guideline for policymakers and social scientists.

**Take-home message:** As shown by this bibliometric analysis, scientific research in Lesbian, Gay, Bisexual, and Transgender (LGBT) had grown gradually but began to gain momentum since 2010, evidencing increased attention to this demographic in the last decade.

**Key words:** Activity index; bibliometric analysis; LGBT; research; topic modeling.

**Cite this paper as:** An D, Batra K. Trends in scientific output on the lesbian, gay, bisexual, and transgender (LGBT) community research: A bibliometric analysis of the literature. *J Health Soc Sci*. 2022;7(1):114-131. Doi: 10.19204/2022/TRND8.

Received: 10 February 2022; Accepted: 15 March 2022; Published: 15 March 2022

---

## INTRODUCTION

The lesbian, gay, bisexual, and transgender (LGBT) community has endured more negative experiences of social inequality compared to the general population [1]. This inequality is often reflected in occupational segregation, healthcare disparities, labor rights, and gender issues. Therefore, it is vital for social scientists and policymakers to learn more about the LGBT population to improve this social and health outcomes. The inequality can basically be attributed to societal rejection of LGBT individuals due to a poor understanding of this population, at least from a scientific perspective; therefore, attempts at solving this inequality require a better understanding of the LGBT community through a lens of science and research [2,3]. By characterizing the current status and historical trend of a scientific field, a bibliometric study can provide guidance to social scientists for policymaking [4,5] and researchers for proposing future research directions [6]. A number of bibliometric studies have been conducted for LGBT research from varying stances, covering the scientific output of gender issues in areas such as family business [7] and child adoption [8], psychological research methods [9], health [10–12] and healthcare-related disparities [10,12,13], and occupational segregation [14]. While previous bibliometric studies in LGBT research have focused primarily on topics such as major influential contributors like countries, journals, institutions, and authors from different perspectives [15], the bibliometric analysis to investigate trends of themes in LGBT research remains elusive until today. Therefore, the present study aims to fill this research gap in LGBT through bibliometric analysis using Python-based methods.

This study set out to (a) evaluate the global research effort in different LGBT topics over time as well as relative scientific contributions from different countries; (b) explore and trace the development of major themes in LGBT research over time. The outcomes of the study should serve as an up-to-date guidance for scholars and policymakers to use when looking ahead to next research or policy agendas regarding the LGBT community.

## METHODS

### *Database and search strategy*

Publication records of LGBT research were collected from the PubMed database and covered a period from the inception to November 21, 2021. The search method and criteria were inspired by previously published bibliometric studies [10,16]. The phrase used for the query was “(LGBT [TIAB] OR Lesbian [TIAB] OR Gay [TIAB] OR Bisexual [TIAB] OR Transgender [TIAB]) AND (has abstract [FILT])” This allowed for the extraction of citations containing the terms “lesbian, gay, bisexual, transgender, or LGBT” in either the title or the abstract. The retrieved dataset included a .csv (comma separated values) metadata file containing information such as the "title", "author", "affiliation", "publication year", "abstract", etc. The search strategy was validated by reviewing the top and bottom 200 abstract records to ensure their relevancy to the LGBT topic.

### *Data tools*

The data analyses were carried out with Python through the Jupyter Notebook environment [17,18]. The GraphPad prism (7.0) programs were used to generate figures for the data obtained from the analyses. Utilization of Python in this bibliometric review was based on consideration of several benefits [19]. First, Python has the ability to handle big metadata from an enormous number of publications. Also, it provides natural language processing (NLP) packages for content analysis through artificial intelligence technology that can transform free text into normalized data suitable for computer processing, a procedure that aids in bibliometric review of large amounts of quantitative data. The main Python packages used for the topic modeling in this study included Pandas, a powerful and flexible open source data analysis tool built on top of the Python programming language [20]; NumPy, a package performing mathematical operations on arrays [21]; Gensim, the richest Python's natural language processing package used for topic modeling [22]; Mallet, a Java-based package for statistical natural language processing [23]; and Spacy, a very useful package for splitting text into words [24].

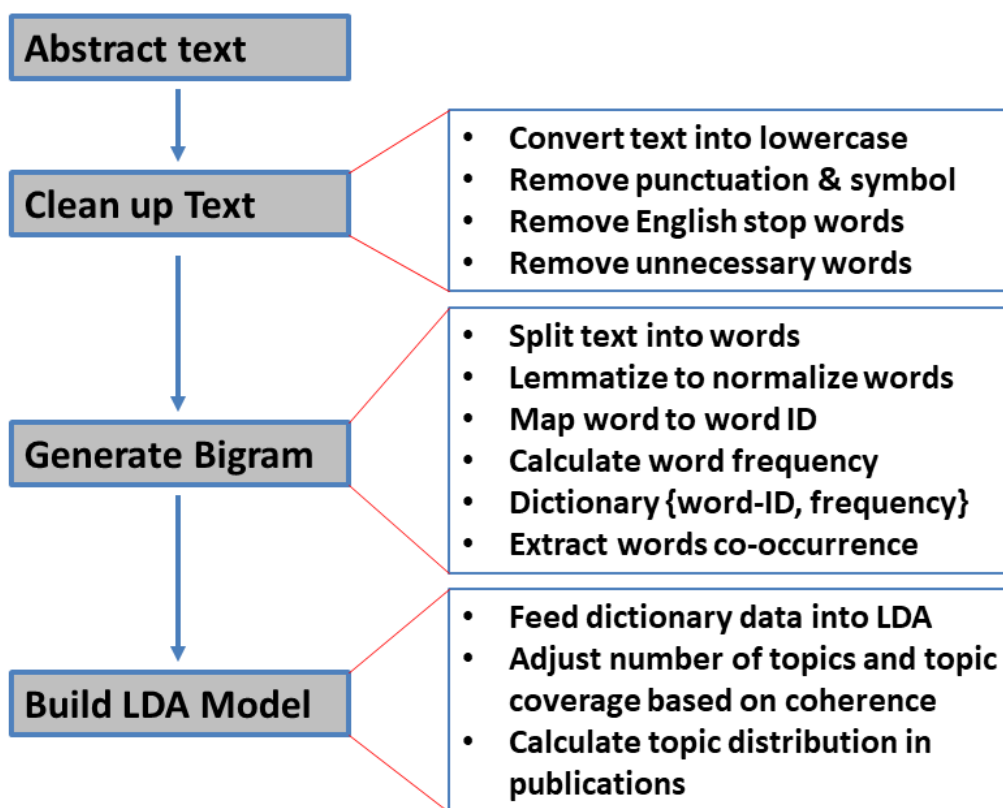
### ***Metadata analysis***

The .csv metadata file retrieved from the PubMed was loaded into Pandas and preprocessed as a data frame. The data frame contained column headers for the metadata fields including "Title", "Author", "Publication Year", "Abstract" and "Affiliation". Publication records with empty or non-English abstracts were removed from the data frame. Duplicate records with the same authors, title, journal, and publication year were eliminated to keep only one in the data frame. In addition to manually reviewing the top and bottom records to validate the search strategy as mentioned above, the field "Abstract" in the data frame was queried with terms "lgbt", "lesbian", "gay", "bisexual" or "transgender" to further verify that each record in the dataset was related to LGBT research before the downstream topic modeling as describe later. The "describe" and "count" methods in Pandas were used for descriptive analysis of the 'Publication Year' field to calculate the number of research articles in LGBT over time. The "for" loop and "dictionary" function were used for iterating over the "Affiliation" field to extract publication information of authorship and countries. The relative research output in LGBT from a country or a certain period of time was evaluated through comparison with the data of publication information in all the research fields that were available in PubMed.

### ***Topic modeling***

The Latent Dirichlet Allocation (LDA) algorithm [25] and the Bigrams model from Gensim, along with Mallet's implementation in Python, were applied for NLP of the "Abstract" field in the dataset to stratify the most popular research topics (Figure 1) [26]. In brief, the abstract contents were first cleaned by removing unnecessary characters such as punctuations, numbers, and section keywords like "background", "method", "result", "conclusion", etc. The cleaned contents were converted to lowercase text and split into lists of words. After removing the English stop words, Bigram model was used to create corpus data of terms co-occurrence frequency for topic modeling. LDA models were trained on the corpus data by feeding with different numbers of topics, and the Spacy package helped normalize and optimize the modeling process to provide better topic segregation. The most favorite model was finally selected through the best topic coherence score corresponding to the modeling outcome which would cover all the major themes without having too many overlaps [22]. Moreover, distribution of an identified topic across the publications was

calculated by measuring the percentage of research articles that covered the topic, to judge how widely the topic was discussed as compared to others. The detailed Python codes used for the data analysis in this study would be available from the corresponding author on reasonable request.



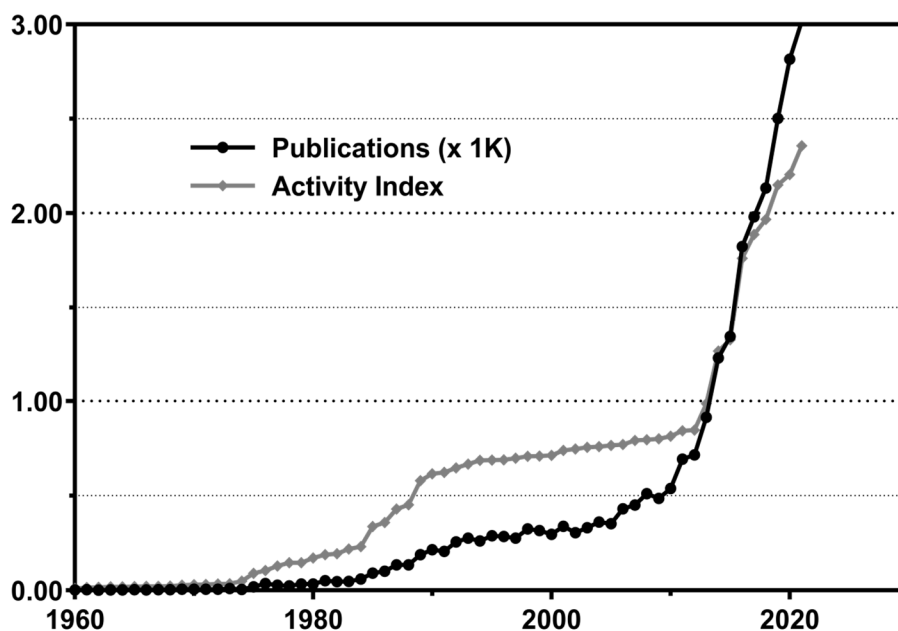
**Figure 1.** The flow chart showing the data processing for LDA topic modeling.

## RESULTS

### *Historical attention to LGBT research*

The LGBT query in the PubMed database resulted in a total of 22,121 records by 11/21/2021. The earliest record could be traced back to year 1892, after which few studies were reported until 1975. The number of publications increased from 4 in 1974 to 18 in 1975, marking a milestone in the record history of LGBT research. Annual growth in research development in the LGBT topic experienced a few waves, with the first one from 1984-1992, the second one from 2005-2010, and the third one from 2010 to present. LGBT research was relatively slow from 1993 to 2004 (Figure 2). The growth of research output has gained momentum since 2010, evidencing a current period of increasing scientific attention being paid to LGBT research. A concept of activity index [27,28] was further adapted to view the historical trend of relative productivity in LGBT research. In this study, the activity index for a given year was calculated as "(LGBT publications in that year / All publications from all the fields in that year) / (All the LGBT publications / All publications in all the fields)". A lower activity index would indicate a relatively less productivity, with a value of 1 corresponding to an average research effort across all the fields. In this study, the activity indices have stayed above 1 since 2014, suggesting that there has been an increasing amount of research activity in LGBT research in recent years compared to other fields (Figure 2). This finding coincides with current trends in the world of

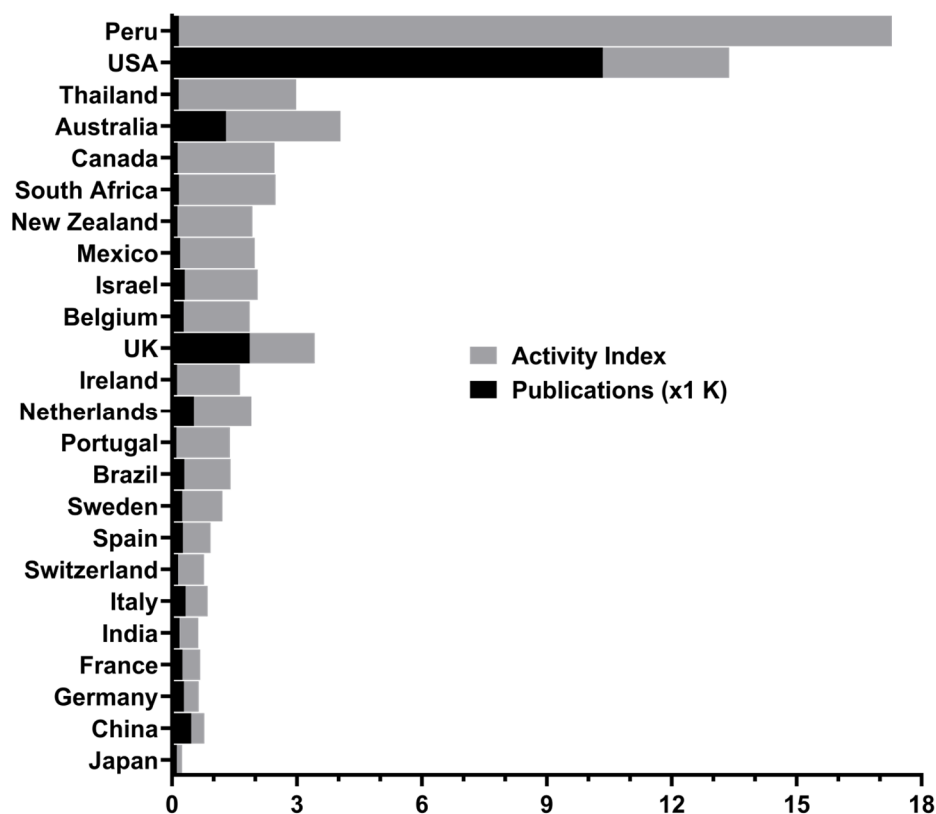
advancing for greater social equality in well-developed regions, which contribute the most to scientific research in general.



**Figure 2.** The trend of scientific publications in LGBT research over time. The number of articles and the activity index in each year indicate the relative scientific attention to LGBT as compared with all the research fields. The yearly activity index in LGBT was calculated as (Publications of LGBT topic in a year / All the publications in that year from all the fields) / (All publications of LGBT topic / All the publications in all the fields). A lower activity index would indicate a relatively less productivity, with a value of 1 corresponding to an average research effort across all the fields.

#### **Regional attention to LGBT research**

The total number of publications from a country might serve as a simple indicator of how much research-related resources the country has put into a specific field. Over 85% of the publications came from the top 20 countries that have contributed the most to LGBT studies, with 9 from Europe (the UK, the Netherlands, Italy, Germany, Belgium, Spain, France, Sweden, and Switzerland), 4 from Asia (China, Israel, India, and Thailand), 2 from North America (the USA and Canada), 3 from South America (Brazil, Mexico, Peru), 1 from Australia (Australia), and 1 from Africa (South Africa), in no particular order. The USA was the most productive country (with > 45% of the total publications), followed by the UK, Canada, Australia, and the Netherlands (Figure 3). However, in another opinion, the number of studies a country conducts should be taken with a grain of salt. With differences in the population, economies, cultures, technologies, etc. between nations, it would make more sense to evaluate a weighted research effort for a given country. Therefore, a concept of activity index would be useful for indicating a country's relative activity in a certain field [27,28]. Herein, activity indices were calculated to compare relative effort in countries that have contributed the greatest number of publications in LGBT research (Figure 3). Peru, the USA, Thailand, Australia, and Canada were the top 5 countries that had relatively allocated more of their scientific effort to LGBT research.



**Figure 3.** The most productive countries and their activity indices in LGBT research. The number of articles contributed by countries and their activity indices indicate their relative effort in LGBT research. The activity index in LGBT research for a country was calculated as (a country's publications in LGBT research / the country's total publications in all fields) / (world output of publications in LGBT research / world total output of publications in all fields). A lower activity index would indicate a relatively less research effort, with an index value of 1 corresponding to the global average.

**Theme trend in LGBT research (topic modeling)**

Topic modeling is an artificial intelligence technology used to identify the main topics of large volumes of text. In this study, LDA with bigrams models were built through Gensim package in Python to identify major topics of LGBT research over time. The dataset of LGBT literature was divided into 6 time periods (1892-1999, 2000-2009, 2010-2014, 2015-2017, 2018-2019, and 2020-2021), with each one containing around 3000-4000 publication records. Optimized LDA with bigrams models corresponding to high coherence scores were built to refine topics and their weighted distribution in each section. The analysis identified 8, 8, 8, 20, 20, and 20 topics from the above timelines, respectively (Supplementary Information Box 1), where each topic was inferred from a combination of 10 keywords and their weight of contribution. For instance (**Table 1**), the "Topic 1: Health - mental health and young adult suicide<sup>0.0467</sup>" in the period of "2020-2021" was represented as "0.058adult + 0.039age + 0.036report + 0.031young + 0.029depression + 0.022high + 0.022symptom + 0.022anxiety + 0.020disorder + 0.018suicide".

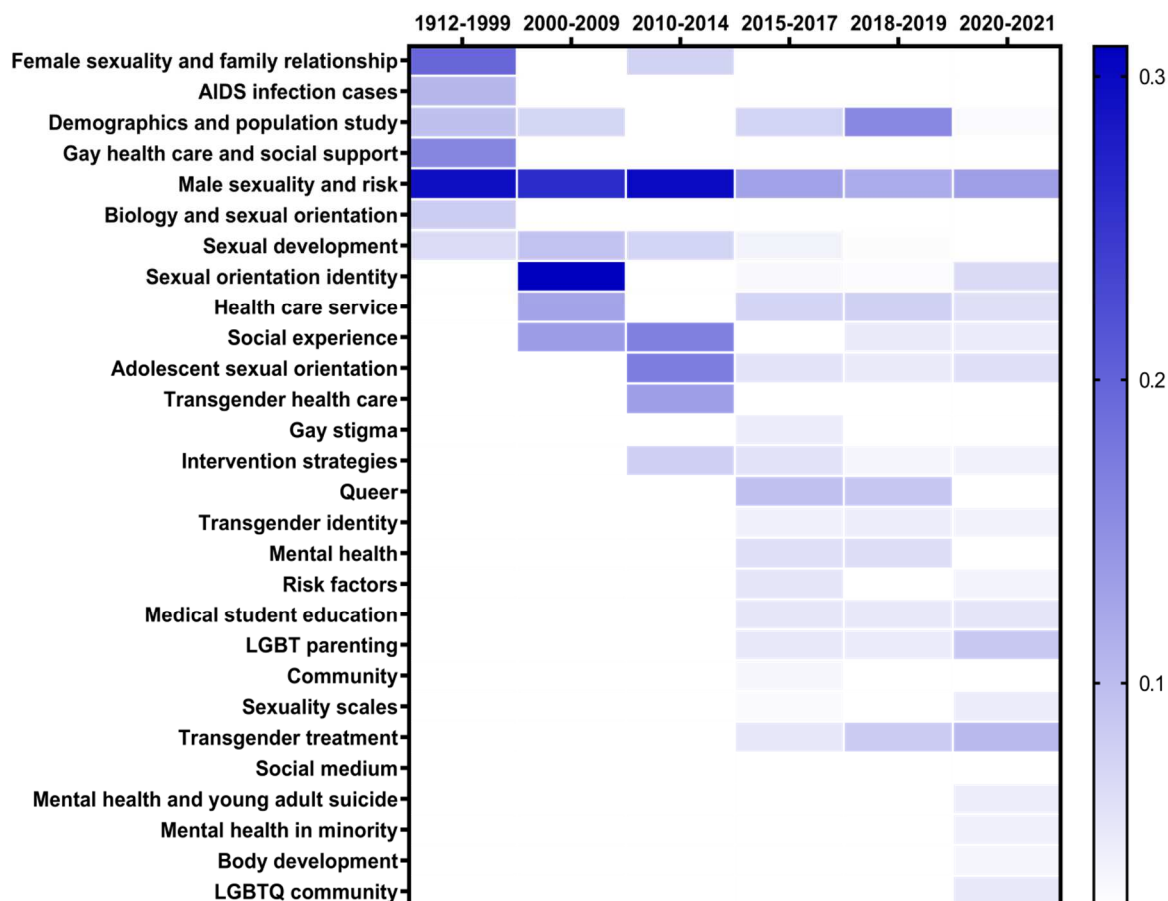
**Table 1.** Major research topics identified from LGBT literature in 2020-2021.

Topic	<b>0: Sexual orientation and gender - adolescent sexual identity</b> <sup>0.0615</sup>
Keywords	<sup>0.145</sup> sexual + <sup>0.069</sup> minority + <sup>0.065</sup> youth + <sup>0.036</sup> identity + <sup>0.030</sup> behavior + <sup>0.025</sup> adolescent + <sup>0.019</sup> victimization + <sup>0.018</sup> school + <sup>0.016</sup> student + <sup>0.015</sup> examine
Topic	<b>1: Health - mental health and young adult suicide</b> <sup>0.0467</sup>
Keywords	<sup>0.058</sup> adult + <sup>0.039</sup> age + <sup>0.036</sup> report + <sup>0.031</sup> young + <sup>0.029</sup> depression + <sup>0.022</sup> high + <sup>0.022</sup> symptom + <sup>0.022</sup> anxiety + <sup>0.020</sup> disorder + <sup>0.018</sup> suicide
Topic	<b>2: Health - transgender treatment</b> <sup>0.0513</sup>
Keywords	<sup>0.048</sup> study + <sup>0.047</sup> treatment + <sup>0.031</sup> include + <sup>0.023</sup> transgender + <sup>0.021</sup> population + <sup>0.021</sup> literature + <sup>0.018</sup> evidence + <sup>0.018</sup> therapy + <sup>0.015</sup> hormone + <sup>0.013</sup> article
Topic	.....
Keywords	.....
Topic	.....
Topic	<b>19: Sexual orientation and gender - male sexuality and risk</b> <sup>0.0584</sup>
Keywords	<sup>0.038</sup> test + <sup>0.032</sup> infection + <sup>0.025</sup> testing + <sup>0.023</sup> high + <sup>0.023</sup> msm + <sup>0.022</sup> population + <sup>0.020</sup> screen + <sup>0.019</sup> positive + <sup>0.019</sup> man + <sup>0.017</sup> prevalence

*Note:* The table shows major research topics of LGBT and their distribution weight across publications in chronological section 2020-2021. Each topic was inferred by a combination of 10 keywords and their weights to the topic. A complete list of research topics identified in this study was summarized in the supplementary information of Box I.

It was interpreted as this topic was addressed by 4.67% (0.0467) of the publications in the period 2020-present and the publications in this topic mainly discussed the issue of mental health and young adult suicide. The top 10 keywords that contributed to this topic were "adult", "age", "report", "young", "depression", "high", "symptom", "anxiety", "disorder" and "suicide" and the weight or importance of the term "suicide" on the topic was 0.018. The topics garnering the most attention in LGBT research over time were "male sexuality and risk", followed by "sexual development", "health care service", "social experience", and "intervention strategies" (Figure 4). The topic of "sexual orientation" has understandably captured attention since 2000 and increased attention has also been paid to the theme "sexual orientation in youth" since 2010. Topics in relation to "transgender" began to get attention in the period 2010-2014. The topics "mental health", "transgender identity", "LGBT parenting", etc. began to receive attention since 2015 and "mental health in youth and minority" since 2020. On the other hand, some topics have faded from the research spotlight over time, such as "female sexuality and family relationship", which was not captured in this study as a major theme after 2015. Evaluating the shift in the major themes of LGBT studies could be used to predict and

forecast future research attention. For example, topics that have taken up significant parts of LGBT research in recent years included "mental health in youth and minority", "intervention strategies", "queer", "risk factors", "medical student education", "body development", "LGBT parenting", etc. These topics would likely continue to be popular in the near future. This information could therefore be used to help researchers and policymakers focus their efforts on nascent topics, or to divert their attention to fields that are less well researched.



**Figure 4.** Weighted distribution of major topics allocated to LGBT research chronologically. To trace the research themes over time, the dataset of LGBT literatures was divided into 6 time periods (1892-1999, 2000-2009, 2010-2014, 2015-2017, 2018-2019, and 2020-present) with each one containing around 3000-4000 publication records. Optimized LDA with bigrams models corresponding to high coherence scores were built to refine topics and their weighted distribution in each section. Weighted distribution of a topic across publications in each time period shown in the heat chart provided information about how widely the topic had been researched during that period.

## DISCUSSION

The present study has provided a comprehensive overview of LGBT research literature and traced scientific attention to LGBT in terms of the evolutionary trend of research effort and themes. To the author's knowledge, this should be the first bibliometric study conducted to chronologically follow the scientific attention and research effort surrounding the LGBT community. The curve of growth trajectory for LGBT literature showed that the scientific reactivity in LGBT research began to gain momentum in 2010 and began to become significantly more popular since 2014 based on the



activity indices. Most of the dominant contributors to scientific effort in LGBT research came from well-developed countries including the USA, the UK, Canada, Australia, the Netherlands, etc. This is most likely because of the greater number of resources available to researchers in these countries, as well as the fact that many scientific journals are home to these countries. Thus, it is worth noting that, although Peru and Thailand did not contribute a high number of publications, they were the top two countries in terms of how much of their scientific effort, relative to their research capabilities, was dedicated to LGBT research. In agreement with the growth trajectory, there was a much greater array of distinct, major topics identified in publications after 2015 (with 8 topics for each time period pre-2015 versus 20 topics for each period post-2015). Together, these results highlight the rising scientific efforts to better understand the LGBT community in an effort to find more progressive solutions to the social inequality they face.

The identified topics in this study reflecting the direction of research effort could be grouped into six general thematic areas: "Sexuality", "Health", "Society", "Culture", "Rights" and "Issues". The first dominant thematic area was "Sexuality" which covered multiple topics identified in the study, including sexual orientation and development, identity, queer, transgender etc. The second prominent thematic area was "Health" which covered aspects of health care, mental health, healthcare service, intervention, medical education, etc. Three topics including AIDS infection disease, risk factors, and gay stigma were identified for the theme "Issues". The theme "Society" included two topics of social experience and social medium. It appears that more future research would be in the direct need for the themes "Culture" and "Rights", which only comprise a single topic of community and parenting, respectively. The major topics and themes from LGBT publications would represent past and current research effort that had been put into to understand and reduce the social inequality. For a representative example, when it was realized that medical providers and health professionals had been often lacking necessary education and trainings required to address sexuality issues [1,29], the topic "medical student education" began to capture scientific attention in the LGBT research since 2015 (Figure 4). Another example would be research topic on transgender. In agreement with the result in this study that the topic of transgender began to dominate since the 2010-2014 chronological section, the number of publications on the topic of transgender had doubled in 2011 after a noticeable growth in the decade of 2000s [10]. On the contrary, topics shown in general social media but irrelevant to the list of keywords distilled in this study (Supplementary Information Box 1) might suggest a lack of sufficient scientific attention to them in LGBT research. For instance, keywords relating to the LGBT rights and health topics like marriage, sport, adoption, religion, and reproduction were not captured from this study; it would recommend that further research attention might be needed on these keywords-relevant topics. Together, the outcomes of the trend in research attention conclude that a comprehensive review is necessary to provide evidence-based guidance for scholars and policymakers to adjust their responses to the next move in LGBT.

In many cases, comprehensive reviews however might confront difficulty when manually dealing with massive citation information, and current study thus consequently provided a basis for digesting big metadata from enormous publications through artificial intelligence technology. Moreover, introduction of artificial intelligence into literature analyses would be a trend in future particularly when publication information grows extremely huge. The data analysis procedure

presented here would provide a good framework for quick review of research effort and trend in other areas of interest as well.

### **Study limitations**

To the best of the author's knowledge, the analysis in the present paper has covered the largest collection of LGBT related research records (n = 22,121) compared to the previous studies. At the same time, it should point out that the current study only queried research articles indexed in PubMed and it may have missed some publications in other databases such as Web of Science and Scopus due to the limit of subscription. However, given the substantial number of publication records included in this study, it should be a pretty confident claim that a large part of the LGBT literature has been covered. Another limitation in the present paper could be that the study only considered the records published in the English language, which might have introduced some bias in the analysis. Future analyses covering publications in other languages from all the major collection databases could be considered. Also, as the focus of the present study was on the trend of research outcome and themes, we did not evaluate the metadata such as journals and authors in this paper. This would recommend a future comprehensive review on the bibliometric indicators in LGBT research. Nevertheless, the bibliometric analysis in this paper allowed an overview of historical and current status in LGBT research and enabled scholars and policymakers to better understand the LGBT from a scientific viewpoint to prioritize the next agenda.

**Author Contributions:** Conceptualization, study design, methodology, formal analysis, writing-original draft, writing- review & editing: DA. Statistical analysis, methodology, and data collection: KB. Resources, and supervision: KB. Writing- review & editing: DA, KB.

**Funding:** None

**Acknowledgments:** None

**Conflicts of Interest:** None

**Data Availability Statement:** Some or all data and models that support the findings of this study are available from the corresponding author upon reasonable request.

**Publisher's Note:** Edizioni FS stays neutral with regard to jurisdictional claims in published maps and institutional affiliation.

## **References**

1. Hughes RL, Damin C, Heiden-Rootes K. Where's the LGBT in integrated care research? A systematic review. *Fam Syst Health*. 2017;35:308–319. Doi: 10.1037/fsh0000290.
2. Saewyc EM. Research on adolescent sexual orientation: development, health disparities, stigma and resilience. *J Res Adolesc*. 2011;21:256–272. Doi: 10.1111/j.1532-7795.2010.00727.x.
3. Bauer GR. Incorporating intersectionality theory into population health research methodology: challenges and the potential to advance health equity. *Soc Sci Med*. 2014;110:10–17.
4. King DA. The scientific impact of nations. *Nature*. 2004;430:311–316.
5. Bar-Ilan J. Informetrics at the beginning of the 21st century - A review. *J Informetr*. 2008;2:1–52.
6. Chang Y-W, Huang M-H, Lin C-W. Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses. *Scientometrics*. 2015;105:2071–2087.

7. Nguyen M-H, Nguyen HTT, Le T-T, Luong A-P, Vuong Q-H. Gender issues in family business research: A bibliometric scoping review. *J Asian Bus Econ Stud*. April 2021. doi: 10.1108/JABES-01-2021-0014.
8. Karthiayani A, Kamthan M. Homosexuality and adoption of children: A Bibliometric Analysis. *Libr Philos Pract*. 2021;5781.
9. Walch SE, Bernal DR, Gibson L, Murray L, Thien S, Steinnecker K. Systematic review of the content and methods of empirical psychological research on LGBTQ and SGM populations in the new millennium. *Psychol Sex Orientat Gend Divers*. 2020;7:433.
10. Sweileh WM. Bibliometric analysis of peer-reviewed literature in transgender health (1900–2017). *BMC Int Health Hum Rights*. 2018;18:1–11.
11. Sweileh WM. Bibliometric analysis of peer-reviewed literature on mental health and well-being of LGBT adolescents. *Ment Health Soc Incl*. April 2022. doi:10.1108/MHSI-02-2022-0013.
12. Barbosa GC, Berigo MR, Assis TR. Health for the LGBT+ population: a bibliometric review. *Rev Psicol Política*. 2020;20:149–164.
13. Ramirez M, Craven HJ. Bibliometric analysis of publications on healthcare disparities among sexual and gender minorities: an exploratory study. Poster presented at the: Bibliometrics & Research Assessment Symposium 2020; October 9, 2020; Online via Webex.
14. Sachdeva L, Bharti K, Maheshwari M. A five-decade review of gender-based occupational segregation: A bibliometric study of influential authors, institutions, and research clusters. *Aust J Career Dev*. 2021;30:117–128.
15. Solkhe A. LGBTQ notion evaluation: A bibliometric analysis and systematic review. *Libr Philos Pract*. 2021:1–38.
16. Kneale D, French R, Henley J, Thomas J. Inequalities and Measurement Issues in Researching Older LGBT PeopleVs Health and Care Needs in the UK: Scoping Review Protocol. London: EPPI-Centre, Institute of Education, University College London; 2018.
17. Van Rossum G, Drake F. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009.
18. Kluyver T, Ragan-Kelley B, Pérez F, et al. Jupyter Notebooks-a publishing format for reproducible computational workflows; 2016.
19. Abd-Alrazaq A, Schneider J, Mifsud B, Alam T, Househ M, Hamdi M, et al. Comprehensive Overview of the COVID-19 Literature: Machine Learning-Based Bibliometric Analysis. *J Med Internet Res*. 2021 Mar 8;23(3):e23703. Doi: 10.2196/23703.
20. McKinney W. Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference: Austin, TX; 2010, pp.51–56.
21. Harris CR, Millman KJ, Van Der Walt SJ, Gommers R, Virtanen R, Cournapeau D, et al. Array programming with NumPy. *Nature*. 2020;585:357–362.
22. Rehurek R, Sojka P. Gensim–python framework for vector space modelling. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic; 2011.
23. McCallum AK. Mallet: A machine learning for language toolkit; 2002.
24. Honnibal M, Montani I. spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing; 2017.
25. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Mach Learn Res*. 2003;3:993–1022.

26. Srinivasa-Desikan B Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras: Packt Publishing Ltd; 2018.
27. de Solla Price DJ. Networks of scientific papers. Science 1965;149:510–515.
28. Velmurugan C. Publication Analysis of Nanotechnology in Global Perspective: a Scientometric Approach. Res J Libr Inf Sci. 2018;2:36–49.
29. Coleman E, Elders J, Satcher D, Shindel A, Parish S, Kenagy G, et al. Summit on medical school education in sexual health: report of an expert consultation. J Sex Med. 2013;10:924–938. Doi: 10.1111/jsm.12142.

© 2022 by the authors. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



### Supplementary Information

**Box I:** Major research topics in LGBT. The box shows major research topics of LGBT and their distribution weight across publications in chronological sections. To trace the research themes over time, the dataset of LGBT literatures was divided into 6 sections (1892-1999, 2000-2009, 2010-2014, 2015-2017, 2018-2019, and 2020-present) with each one containing around 3,000-4,000 publication records. Optimized LDA with bigrams models corresponding to high coherence scores were built to refine topics and their weighted distribution in each section. Each topic was inferred by a combination of 10 keywords and their weights to the topic.

#### *Section 1912-1999*

**Topic 0:** Sexual orientation and gender - female sexuality and family relationship <sup>0.1952</sup>

0.051lesbian + 0.044woman + 0.028heterosexual + 0.020sexual + 0.018orientation + 0.017homosexual + 0.015child + 0.015relationship + 0.015family + 0.014identity

**Topic 1:** Sexual orientation and gender - aids infection cases <sup>0.1074</sup>

0.052aid + 0.032year + 0.026increase + 0.025man + 0.025case + 0.023time + 0.021rate + 0.020bisexual + 0.018age + 0.015infection

**Topic 2:** Health - gay health care and social support <sup>0.1606</sup>

0.080gay + 0.032health + 0.026social + 0.021community + 0.018support + 0.018care + 0.013provide + 0.011people + 0.010program + 0.009base

**Topic 3:** Sexual orientation and gender - male sexuality and risk <sup>0.1582</sup>

0.083man + 0.059risk + 0.058sexual + 0.042sex + 0.036partner + 0.033behavior + 0.030report + 0.021high + 0.017bisexual + 0.015condom

**Topic 4:** Sexual orientation and gender - sexuality and infection disease <sup>0.1347</sup>

0.064patient + 0.044infection + 0.027homosexual + 0.024positive + 0.022disease + 0.018infect + 0.017test + 0.015bisexual + 0.013antibody + 0.012heterosexual

**Topic 5:** Sexual orientation and gender - demographics and population study <sup>0.0962</sup>

0.034man + 0.024subject + 0.022level + 0.021relate + 0.020group + 0.019significant + 0.017significantly + 0.016measure + 0.016control + 0.016associate

**Topic 6:** Sexual orientation and gender - biology and sexual orientation <sup>0.0815</sup>

0.020population + 0.019bisexual + 0.011cell + 0.010study + 0.008show + 0.007site + 0.007specie + 0.007suggest + 0.007sequence + 0.007single

**Topic 7:** Sexual orientation and gender - sexual development 0.0661

0.062male + 0.036female + 0.032group + 0.013development + 0.009adult + 0.008response + 0.008early + 0.007give + 0.007week + 0.007describe

#### Section 2000-2009

**Topic 0:** Sexual orientation and gender - male sexuality and risk 0.1499

0.082sex + 0.049risk + 0.041sexual + 0.040partner + 0.035behavior + 0.026report + 0.020msm + 0.017drug + 0.015prevention + 0.013associate

**Topic 1:** Sexual orientation and gender - sexual identity of lesbian and gay 0.1832

0.019community + 0.014lesbian + 0.014gay + 0.013identity + 0.012article + 0.009paper + 0.009sexuality + 0.009process + 0.008focus + 0.008context

**Topic 2:** Health - health care service 0.1278

0.060health + 0.024care + 0.016people + 0.016include + 0.016provide + 0.015issue + 0.014service + 0.012transgender + 0.011base + 0.010mental

**Topic 3:** Sexual orientation and gender - sexual development 0.0929

0.025male + 0.016female + 0.011bisexual + 0.010model + 0.007development + 0.006find + 0.006stage + 0.006form + 0.006show + 0.005potential

**Topic 4:** Sexual orientation and gender - male sexuality and risk 0.1114

0.141man + 0.027infection + 0.025gay + 0.023positive + 0.019increase + 0.018risk + 0.017patient + 0.016year + 0.015test + 0.014case

**Topic 5:** Sexual orientation and gender - sexual orientation identity 0.1266

0.076sexual + 0.068woman + 0.054bisexual + 0.045heterosexual + 0.030orientation + 0.022lesbian + 0.022male + 0.019difference + 0.017report + 0.017homosexual

**Topic 6:** Sexual orientation and gender - demographics and population study 0.0724

0.033group + 0.023high + 0.021level + 0.018factor + 0.016individual + 0.016treatment + 0.016low + 0.015effect + 0.015associate + 0.013study

**Topic 7:** Society - social experience and family support 0.1359

0.081gay+0.033lesbian+0.031relationship+0.029social+0.027experience+0.023family+0.022support+0.017participant+0.016child+ 0.013life

#### Section 2010-2014

**Topic 0:** Health - community intervention strategy 0.0801

0.027intervention + 0.024base + 0.021participant + 0.021population + 0.017increase + 0.017prevention + 0.016group + 0.013community + 0.012strategy + 0.011include

**Topic 1:** Sexual orientation and gender - sexual development 0.0741

0.034male + 0.031female + 0.011bisexual + 0.010development + 0.010case + 0.008early + 0.008change + 0.006show + 0.005control + 0.005flower

**Topic 2:** Sexual orientation and gender - male sexuality and risk 0.1932

0.071sex + 0.066man + 0.040partner + 0.039risk + 0.033msm + 0.024report + 0.019high + 0.018test + 0.018infection + 0.017behavior

**Topic 3:** Society - social experience 0.1669

0.032social + 0.025experience + 0.021identity + 0.019community + 0.014article + 0.012stigma + 0.012explore + 0.011work + 0.011focus + 0.010interview

**Topic 4:** Sexual orientation and gender - adolescent sexual orientation <sup>0.1702</sup>

0.060sexual + 0.025bisexual + 0.025minority + 0.022health + 0.019risk + 0.019high + 0.019orientation + 0.018youth + 0.016age + 0.016heterosexual

**Topic 5:** Sexual orientation and gender - female sexuality and family relationship <sup>0.0765</sup>

0.071woman + 0.047lesbian + 0.037relationship + 0.029heterosexual + 0.019family + 0.017couple + 0.016study + 0.016difference + 0.015child + 0.015group

**Topic 6:** Health - transgender health care <sup>0.1337</sup>

0.063health + 0.036transgender + 0.032care + 0.021people + 0.019gender + 0.017patient + 0.017service + 0.017treatment + 0.016include + 0.013individual

**Topic 7:** Sexual orientation and gender - male sexuality <sup>0.1052</sup>

0.096gay + 0.060man + 0.060sexual + 0.027bisexual + 0.023participant + 0.013orientation + 0.012negative + 0.012measure + 0.011finding + 0.011gender

### Section 2015-2017

**Topic 0:** Issues - gay stigma <sup>0.047</sup>

0.048gay + 0.041stigma + 0.033perceive + 0.027relate + 0.024level + 0.023negative + 0.020perception + 0.018attitude + 0.016disclosure + 0.015examine

**Topic 1:** Sexual orientation and gender - demographics and population study <sup>0.0386</sup>

0.053study + 0.053population + 0.043include + 0.031identify + 0.023evidence + 0.022literature + 0.021people + 0.017base + 0.015key + 0.015country

**Topic 2:** Sexual orientation and gender - demographics and population study <sup>0.0195</sup>

0.113participant + 0.057report + 0.047survey + 0.036sample + 0.034identify + 0.034online + 0.026recruit + 0.022complete + 0.019status + 0.019positive

**Topic 3:** Sexual orientation and gender - male sexuality and HIV risk <sup>0.0692</sup>

0.075msm + 0.047test + 0.044man + 0.042infection + 0.033testing + 0.031high + 0.027risk + 0.018hiv + 0.017prevalence + 0.016diagnosis

**Topic 4:** Health - intervention and prevention strategies <sup>0.0581</sup>

0.058intervention + 0.039prep + 0.031prevention + 0.016strategy + 0.016base + 0.016reduce + 0.013effective + 0.010improve + 0.010program + 0.010target

**Topic 5:** Sexual orientation and gender - male sexuality and risk <sup>0.0608</sup>

0.180sex + 0.083man + 0.079partner + 0.050sexual + 0.027behavior + 0.027risk + 0.024relationship + 0.022report + 0.020condom + 0.017male

**Topic 6:** Sexual orientation and gender - demographics and population study <sup>0.0168</sup>

0.063age + 0.057adult + 0.041rate + 0.038group + 0.035compare + 0.029year + 0.028time + 0.027difference + 0.026population + 0.020prevalence

**Topic 7:** Health - health care service <sup>0.0744</sup>

0.197health + 0.103care + 0.050service + 0.033provider + 0.029access + 0.024healthcare + 0.018patient + 0.018mental + 0.016experience + 0.015barrier

**Topic 8:** Sexual orientation and gender - queer <sup>0.0964</sup>

0.016context + 0.016queer + 0.015article + 0.015sexuality + 0.015social + 0.014work + 0.013community + 0.013approach + 0.013interview + 0.012experience

**Topic 9:** Sexual orientation and gender - transgender identity <sup>0.0428</sup>

0.181transgender + 0.140gender + 0.087individual + 0.058identity + 0.055people + 0.028experience + 0.022person + 0.016transition + 0.015tran + 0.014cisgender

**Topic 10:** Health - mental health 0.0623

0.043support + 0.039mental + 0.033health + 0.028depression + 0.028social + 0.027stress + 0.025discrimination + 0.024model + 0.022relationship + 0.022associate

**Topic 11:** Sexual orientation and gender - sexual development 0.0408

0.052male + 0.040female + 0.011show + 0.007development + 0.007bisexual + 0.006find + 0.006potential + 0.006specie + 0.006structure + 0.006pattern

**Topic 12:** Health - risk factors 0.0554

0.073risk + 0.059high + 0.045associate + 0.036factor + 0.029report + 0.028substance + 0.025aor + 0.024drug + 0.021alcohol + 0.020odd

**Topic 13:** Sexual orientation and gender - sexual orientation identity 0.0334

0.167woman + 0.139man + 0.116gay + 0.109bisexual + 0.069lesbian + 0.053heterosexual + 0.019difference + 0.013finding + 0.013compare + 0.010identify

**Topic 14:** Health - surgery treatment 0.0534

0.054patient + 0.036treatment + 0.020female + 0.017male + 0.013transgender + 0.012level + 0.012surgery + 0.011therapy + 0.011case + 0.011hormone

**Topic 15:** Society - medical student education 0.0544

0.035student + 0.029knowledge + 0.024education + 0.020practice + 0.019school + 0.018program + 0.018medical + 0.017issue + 0.017lgbtq + 0.015provide

**Topic 16:** Sexual orientation and gender - adolescent sexual orientation 0.0576

0.249sexual + 0.071orientation + 0.066minority + 0.053youth + 0.034heterosexual + 0.027bisexual + 0.025adolescent + 0.025identity + 0.025experience + 0.022behavior

**Topic 17:** Society - parenting 0.0514

0.043family + 0.029couple + 0.028child + 0.019relationship + 0.019parent + 0.014support + 0.013important + 0.013concern + 0.011client + 0.010legal

**Topic 18:** Culture - community 0.0361

0.056young + 0.054social + 0.038community + 0.027black + 0.023group + 0.020live + 0.018network + 0.018base + 0.013medium + 0.013examine

**Topic 19:** Sexual orientation and gender - sexuality scales 0.0314

0.042measure + 0.023scale + 0.023life + 0.020score + 0.019analysis + 0.018satisfaction + 0.016assess + 0.015show + 0.015group + 0.014quality

### Section 2018-2019

**Topic 0:** Sexual orientation and gender - queer 0.0897

0.020queer + 0.013sexuality + 0.012article + 0.011policy + 0.010cultural + 0.008practice + 0.008work + 0.008identity + 0.008legal + 0.008change

**Topic 1:** Sexual orientation and gender - male sexuality and risk 0.0507

0.141sex + 0.055man + 0.054partner + 0.043report + 0.040risk + 0.036sexual + 0.030drug + 0.023msm + 0.020associate + 0.016behavior

**Topic 2:** Sexual orientation and gender - demographics and population study 0.0359

0.025effect + 0.025individual + 0.024model + 0.018show + 0.014find + 0.014response + 0.013measure + 0.013perceive + 0.013level + 0.012role

**Topic 3:** Health - mental health and anti-LGBT issues 0.0646

0.057health + 0.051mental + 0.041experience + 0.026depression + 0.023stress + 0.023violence + 0.021physical + 0.020discrimination + 0.020life + 0.019victimization

**Topic 4:** Sexual orientation and gender - demographics and population study 0.0441

0.086prep + 0.076participant + 0.024follow + 0.022month + 0.017increase + 0.013cohort + 0.013baseline + 0.012pre\_exposure + 0.012adherence + 0.012time

**Topic 5:** Sexual orientation and gender - sexual development 0.0287

0.085male + 0.068female + 0.014sex + 0.011case + 0.010development + 0.009analysis + 0.007size + 0.006cluster + 0.006obtain + 0.006difference

**Topic 6:** Health - transgender treatment 0.0835

0.059patient + 0.024treatment + 0.019transgender + 0.017surgery + 0.016hormone + 0.012testosterone + 0.012procedure + 0.012therapy + 0.012outcome + 0.011undergo

**Topic 7:** Society - medical student education 0.0515

0.035student + 0.035knowledge + 0.028education + 0.023transgender + 0.020survey + 0.019medical + 0.019patient + 0.018practice + 0.017training + 0.017lgbtq

**Topic 8:** Health - community intervention 0.0369

0.066intervention + 0.028community + 0.022base + 0.021treatment + 0.021population + 0.020strategy + 0.020prevention + 0.017reduce + 0.015program + 0.014key

**Topic 9:** Sexual orientation and gender - adolescent sexual orientation 0.0497

0.232sexual + 0.087minority + 0.064youth + 0.059orientation + 0.040identity + 0.035heterosexual + 0.029bisexual + 0.025behavior + 0.022adolescent + 0.019disparity

**Topic 10:** Health - health care service 0.0784

0.160health + 0.126care + 0.049service + 0.040provider + 0.033access + 0.031healthcare + 0.026barrier + 0.019patient + 0.016provide + 0.013include

**Topic 11:** Sexual orientation and gender - male sexuality and risk 0.0682

0.042test + 0.040msm + 0.039infection + 0.029testing + 0.029man + 0.022high + 0.021increase + 0.021positive + 0.017risk + 0.016prevalence

**Topic 12:** Sexual orientation and gender - transgender identity 0.0464

0.191gender + 0.171transgender + 0.063individual + 0.045identity + 0.036people + 0.028cisgender + 0.021tran + 0.017transition + 0.016person + 0.015treatment

**Topic 13:** Sexual orientation and gender - demographics and population study 0.04

0.049study + 0.029population + 0.029include + 0.025literature + 0.021scale + 0.019article + 0.018evidence + 0.017identify + 0.017measure + 0.014quality

**Topic 14:** Society - parenting 0.0489

0.048family + 0.046relationship + 0.040child + 0.026parent + 0.024couple + 0.016disclosure + 0.015role + 0.015desire + 0.014important + 0.012gay

**Topic 15:** Society - social medium 0.0269

0.034online + 0.029information + 0.022base + 0.022identify + 0.021social + 0.019survey + 0.016medium + 0.015sample + 0.014public + 0.012recruit

**Topic 16:** Health - risk factors 0.0384

0.063associate + 0.059factor + 0.058high + 0.055risk + 0.037low + 0.030substance + 0.028level + 0.020sample + 0.019association + 0.018analysis

**Topic 17:** Society - social experience 0.0497



0.064experience + 0.056social + 0.040support + 0.032stigma + 0.032participant + 0.028community + 0.026interview + 0.019qualitative + 0.018explore + 0.017live

**Topic 18:** Sexual orientation and gender - sexual orientation identity 0.0297

0.149woman + 0.144man + 0.120gay + 0.093bisexual + 0.038heterosexual + 0.037lesbian + 0.022black + 0.020group + 0.015compare + 0.015examine

**Topic 19:** Sexual orientation and gender - demographics and population study 0.0397

0.054adult + 0.053age + 0.053compare + 0.035group + 0.034report + 0.030high + 0.024population + 0.021prevalence + 0.020young + 0.020rate

### Section 2020-2021

**Topic 0:** Sexual orientation and gender - adolescent sexual identity 0.0615

0.145sexual + 0.069minority + 0.065youth + 0.036identity + 0.030behavior + 0.025adolescent + 0.019victimization + 0.018school + 0.016student + 0.015examine

**Topic 1:** Health - mental health and young adult suicide 0.0467

0.058adult + 0.039age + 0.036report + 0.031young + 0.029depression + 0.022high + 0.022symptom + 0.022anxiety + 0.020disorder + 0.018suicide

**Topic 2:** Health - transgender treatment 0.0513

0.048study + 0.047treatment + 0.031include + 0.023transgender + 0.021population + 0.021literature + 0.018evidence + 0.018therapy + 0.015hormone + 0.013article

**Topic 3:** Sexual orientation and gender - demographics and population study 0.0322

0.040year + 0.029increase + 0.027time + 0.025participant + 0.025change + 0.025follow + 0.025month + 0.019age + 0.018cohort + 0.016compare

**Topic 4:** Sexual orientation and gender - male sexuality and risk 0.0754

0.090sex + 0.084prep + 0.056man + 0.044partner + 0.033risk + 0.030msm + 0.026report + 0.024prevention + 0.015participant + 0.014increase

**Topic 5:** Society and rights - parenting 0.0871

0.034family + 0.025child + 0.020parent + 0.013queer + 0.012couple + 0.010understand + 0.010context + 0.009article + 0.009discuss + 0.008explore

**Topic 6:** Society - medical student education 0.0556

0.031knowledge + 0.029student + 0.028education + 0.024medical + 0.020survey + 0.018training + 0.017practice + 0.015patient + 0.014program + 0.013regard

**Topic 7:** Health - transgender surgery 0.0529

0.107patient + 0.027surgery + 0.017undergo + 0.016transgender + 0.016outcome + 0.014procedure + 0.014surgical + 0.014affirm + 0.012case + 0.012perform

**Topic 8:** Health - mental health in minority 0.0431

0.182health + 0.071mental + 0.031stress + 0.030minority + 0.029population + 0.023outcome + 0.020relate + 0.017disparity + 0.016individual + 0.016physical

**Topic 9:** Health - online intervention strategies 0.0414

0.056intervention + 0.053participant + 0.034base + 0.018information + 0.018online + 0.012include + 0.011recruit + 0.011young + 0.011design + 0.011conduct

**Topic 10:** Sexual orientation and gender - sexual orientation identity 0.0392

0.159woman + 0.122man + 0.110sexual + 0.098gay + 0.081bisexual + 0.053heterosexual + 0.037orientation + 0.030lesbian + 0.019difference + 0.017compare

**Topic 11:** Sexual orientation and gender - body development <sup>0.0375</sup>

0.009development + 0.009body + 0.008pregnancy + 0.008show + 0.008analysis + 0.007effect + 0.007increase + 0.007potential + 0.007human + 0.007level

**Topic 12:** Sexual orientation and gender - transgender identity <sup>0.0412</sup>

0.203gender + 0.182transgender + 0.060people + 0.046identity + 0.043cisgender + 0.041individual + 0.031tran + 0.024affirm + 0.023diverse + 0.020birth

**Topic 13:** Culture - lgbtq community and policy <sup>0.0515</sup>

0.047community + 0.028lgbtq + 0.022policy + 0.018qualitative + 0.017people + 0.016interview + 0.015focus + 0.014approach + 0.014theme + 0.013identify

**Topic 14:** Society - social experience <sup>0.0488</sup>

0.086experience + 0.071social + 0.046support + 0.034stigma + 0.034discrimination + 0.033violence + 0.023participant + 0.018relationship + 0.017relate + 0.016live

**Topic 15:** Health - health care service <sup>0.0619</sup>

0.131care + 0.077health + 0.057service + 0.041provider + 0.039access + 0.032healthcare + 0.031barrier + 0.023provide + 0.014include + 0.014primary

**Topic 16:** Health - risk factors <sup>0.0392</sup>

0.067risk + 0.058high + 0.048factor + 0.047associate + 0.027substance + 0.027aor + 0.025low + 0.024report + 0.022odd + 0.022prevalence

**Topic 17:** Sexual orientation and gender - sexuality scales <sup>0.0474</sup>

0.032measure + 0.025score + 0.023sample + 0.022scale + 0.022level + 0.019individual + 0.019relationship + 0.018analysis + 0.017effect + 0.016assess

**Topic 18:** Sexual orientation and gender - sexual orientation identity <sup>0.0279</sup>

0.067male + 0.064female + 0.062sex + 0.046group + 0.027identify + 0.016difference + 0.015base + 0.013finding + 0.012find + 0.011pattern

**Topic 19:** Sexual orientation and gender - male sexuality and risk <sup>0.0584</sup>

0.038test + 0.032infection + 0.025testing + 0.023high + 0.023msm + 0.022population + 0.020screen + 0.019positive + 0.019man + 0.017prevalence